DOCUMENT RESUME

ED 312 211 SO 020 341

AUTHOR Fraenkel, Jack R.; Wallen, Norman E.
TITLE Toward Improving Research in Social Studies Education. SSEC Monograph Series.
INSTITUTION Social Science Education Consortium, Inc., Boulder, Colo.
REPORT NO ISBN-0-89994-325-X
PUB DATE 88
NOTE 68p.
AVAILABLE FROM Social Science Education Consortium, Inc., 855 Broadway, Boulder, CO 80302.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Classroom Research; *Educational Research; *Evaluation; Research Design; Research Methodology; Research Problems; *Social Studies

ABSTRACT

Social studies research has been criticized for sampling bias, inappropriate methodologies, incorrect or inappropriate use of statistics, weak or ill-defined treatments, and lack of replication and/or longitudinal follow-up. In an effort to ascertain whether past criticisms were true of current research as well, a review was conducted of 118 studies published in "Theory and Research in Social Education" (TRSE), the "Journal of Social Studies Research" (JSSR), and the research section of "Social Education" (SE) for the years 1979-1986. This monograph, the first in a series designed to provide "cutting edge" information to the social studies profession, presents the results of this investigation. Chapter 1 critiques the 118 studies. Chapter 2 offers some observations, based on the analysis in chapter 1, about the nature of current social studies research. Chapter 3 discusses how the quality of social studies research might be improved. The remarks in this chapter are directed to three groups of social studies educators: (1) professors who direct master's theses or doctoral dissertations, but who do not teach courses in educational research, (2) graduate students who intend to do research, and (3) classroom teachers who have an interest in research. Chapter 4 contains an in-depth evaluation of a single study using the same criteria discussed in chapter 1. Chapter 5 presents some ideas about how classroom teachers of social studies might become more involved in research in their classrooms and schools. Chapter 6 lists the studies reviewed. A 99-item bibliography is included. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312211

Toward Improving Research in Social Studies Education

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JAMES R. GIESE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SO 020 341

Jack R. Fraenkel and Norman E. Wallen

SSEC Monograph Series
Social Science Education Consortium
Boulder, Colorado

BEST COPY AVAILABLE

TOWARD IMPROVING RESEARCH IN SOCIAL STUDIES EDUCATION

by

Jack R. Fraenkel and Norman E. Wallen
San Francisco State University

SSEC Monograph Series

Social Science Education Consortium
Boulder, Colorado

1988

ORDERING INFORMATION

This publication is available from:
Social Science Education Consortium, Inc.
855 Broadway
Boulder, Colorado 80302

ISBN 0-89994-325-X

© 1988 Social Science Education Consortium, Inc. Reproduction by any means, electronic or mechanical, is prohibited without written permission from the publisher.

ACKNOWLEDGMENTS

Portions of this monograph were earlier published as articles in *Theory and Research in Social Education*. See Jack R. Fraenkel, Toward improving research in social studies education, *Theory and Research in Social Education*, 15 (3); also see Norman E. Wallen and Jack R. Fraenkel, An analysis of social studies research over an eight year period, *Theory and Research in Social Education*, 16 (1). Reprinted by permission of the College and University Faculty Assembly, National Council for the Social Studies.

The authors would like to extend their appreciation to Fred Newmann of the University of Wisconsin for his thoughtful reviews and comments on earlier drafts of the manuscript. The efforts of James E. Davis in coordinating the monograph series are also appreciated.

ABOUT THE AUTHORS

Jack R. Fraenkel and Norman E. Wallen are both professors of interdisciplinary studies in education and part of the Research and Development Center at San Francisco State University. They specialize in research conceptualization and design.

CONTENTS

Foreword iv

Introduction 1

1. **An Assessment of Current Social Studies Research** 3
   - Overview of the Study 3
   - Procedures 5
   - Results of Analysis 6

2. **Some Summary Observations About Social Studies Research** 23
   - Methodology 23
   - Focus/Clarity 23
   - Sample 24
   - Replication 24
   - Internal Validity 24
   - Reliability and Validity of Instruments 24
   - External Validity 24
   - Theory 24
   - Investigators 24
   - Data Sources 24
   - Topics Investigated 25

3. **Suggestions for Improving the Quality of Social Studies Research** 29
   - Improving Experimental Research 29
   - Improving Survey Research 31
   - Improving Correlational or Causal-Comparative Research 32
   - Improving Ethnographic Research 33
   - Some Ideas for Improving Research in General 34

4. **A Detailed Analysis of a Sample Study** 37
   - The Study 37
   - The Analysis 45

5. **Suggestions for Classroom Research in the Social Studies** 51
   - Experimental Research 51
   - Survey Research 52
   - Content Analysis 53
   - Correlational Research 53
   - Causal-Comparative Research 54
   - Ethnographic and Case Study Research 54
   - A Final Word 55

6. **Studies Reviewed** 57

References 63

FOREWORD

In late 1986, the Social Science Education Consortium Board of Directors launched the SSEC Monograph Program. Recognizing that sound, scholarly work often goes unnoticed, the Board decided to take the initiative in continuing the mission for which the SSEC has become known—offering "cutting edge" information to the social studies profession. Thus, the purpose of the SSEC Monograph Program is to publish scholarly monographs in the field of social studies/social science education that make a significant contribution to the profession. It is with this purpose in mind that we offer *Toward Improving Research in Social Studies Education* as the first SSEC monograph.

In 1985 we were painfully reminded by several scholars of a host of defects in social studies research—lack of replicability, lack of innovative methodology, lack of external and internal validity, and inappropriate application of statistical techniques to name a few. Some scholars even used the terms "trivial" and "mindless" to characterize the research in our field (see Stanley 1985). Jack Fraenkel and Norm Wallen took the criticism of research in social studies seriously. They decided to examine systematically actual studies. In preparing this monograph, they used rigorous criteria to analyze 118 studies published in three major sources—*Theory and Research in Social Studies* (TRSE), the *Journal of Social Studies Research* (JSSR), and the research section of *Social Education* (SE)—for the years 1979 through 1986. They take their analysis a step farther, providing helpful suggestions for improvement to professors, graduate students, and classroom teachers who may be planning to conduct research. To bring their analytical criteria to an operational level, they critique one study in detail, illustrating both its strengths and weaknesses. Finally, they offer practical suggestions to classroom teachers who may become involved in school and classroom research.

This is not a volume that should be purchased by only a few scholars to be put in the research section of a personal library. The monograph is important to any social studies professional who wants to avoid the research mistakes of the past, or who wishes to critique a research study or proposal. The standards for judging research are an important guide for any committed social studies educator, whether or not they are engaged in research. *Toward Improving Research in Social Studies Education* is "must" reading for all social studies educators.

James E. Davis, Chair
SSEC Board of Directors
Publications Committee

Do you have an unpublished study or treatise on social studies/social science education that you would like to have considered for the new SSEC monograph series? If so, send a proposal and outline to:

SSEC Board of Directors
c/o Marcia Hutson
855 Broadway
Boulder, CO 80302

INTRODUCTION

Criticisms of the nature and quality of educational research in general continue to appear in the professional literature. More and more frequently, one sees arguments or proposals for changing not only the nature of research, but also the standards by which it is judged. There have been arguments to move toward qualitative (as opposed to quantitative) analyses, to integrate quantitative and qualitative methods of inquiry, even to consider and develop new methods of inquiry altogether (Allender 1986). Researchers have been urged to place less emphasis on external validity (Mook 1983), to decrease their use of inferential statistics (Carver 1978), to concentrate on common-sense interpretations and replication to promote understanding (Stake 1978), to consider introspection and speculation as valid scientific methods (Bakan 1975), to conduct unrationalized (i.e., unplanned) studies (Larkins and Puckett 1983), and even to consider art as a model for scientific investigation (Eisner 1981).

Research in social studies education has not escaped these criticisms and suggestions. Social studies research has been criticized for sampling bias, inappropriate methodologies, incorrect or inappropriate use of statistics, weak or ill-defined treatments, and lack of replication and/or longitudinal follow-up. Many social studies research questions are said to be trivial. Control and experimental groups are seldom equivalent. Hawthorne or John Henry effects contaminate findings. Aptitude-treatment interactions are almost totally ignored. Instruments are poorly designed, frequently lacking validity or reliability. The durability of effects, when any are detected, is almost never assessed. Statistical procedures are frequently inappropriate. Legitimate generalizability is almost nonexistent (e.g., see Cornbleth 1982; Fraenkel 1987; Larkins and McKinney 1980; Leming 1985; Martorella 1977; Nelson and Shaver 1985; Newmann 1985; Shaver 1979b; Shaver and Norton 1980; Wallen 1983; Wallen and Fraenkel 1988).

In the late 1970s, Shaver and Norton (1980) reported that only a small percentage of 53 articles in two social studies journals involved random sampling or assignment, replication of previous work, or limited their conclusions due to shortcomings in accessible populations and samples. Intrigued by their findings, we decided to investigate whether current social studies research efforts continue to suffer from these (or other) faults. We wanted to see if past criticisms were true of current research efforts as well.

Accordingly, we reviewed the research (with certain exceptions) reported in *Theory and Research in Social Education* (TRSE), *The Journal of Social Studies Research* (JSSR), and the research section of *Social Education* (SE) for the years 1979-1986. We wanted to look at a number of characteristics in addition to those which Shaver and Norton studied, however. This monograph presents the results of our work.

The monograph contains six chapters. In the first, we critique all of the empirical studies published in TRSE, JSSR, and the research section of SE between 1979 and 1986. In Chapter 2, we offer some observations, based on the analysis in Chapter 1, about the nature of current social studies research.

In Chapter 3, we offer some ideas about how the quality of social studies research might be improved. We direct the remarks in this section to three distinct groups of social studies educators: (1) professors who direct master's theses or doctoral dissertations, but who do not teach courses in educational research, (2) graduate students who intend to do research, and (3) classroom teachers who have an interest in research.

In Chapter 4, we evaluate a single study in depth, using the same criteria discussed in Chapter 1. We analyze both the weaknesses and strengths of this study in order to illustrate not only those procedures we believe researchers should avoid, but also those they should employ in social studies research.

In Chapter 5, we offer some ideas about how classroom teachers of social studies might become more involved in research in their classrooms and schools. Chapter 6 lists the studies reviewed.

CHAPTER 1
AN ASSESSMENT OF CURRENT SOCIAL STUDIES RESEARCH

Overview of the Study

We reviewed all of the research published from 1979 through 1986 in *Theory and Research in Social Education* (volumes 7 through 15), *The Journal of Social Studies Research* (volumes 3 through 10), and the research section of *Social Education* (volumes 43 through 49), with certain exceptions (as noted below). The instrument shown in Figure 1 (see page 4) was used to analyze and evaluate the studies.

Articles falling in one or more of the following categories were not analyzed:

- Arguments or position papers, in which the author(s) argued that a particular position or program of some sort should be adopted or considered by the social studies profession.
- Historical studies, in which the author(s) described, reviewed, and/or analyzed some aspect of social studies education in the past.
- Content analyses, in which the author(s) analyzed the contents of textbooks or other types of social studies documents.
- Philosophical inquiries, in which the author(s) presented rationale statements of some sort or delved into the meaning of various terms used by social studies professionals.
- Methodological proposals, in which the author(s) proposed that a certain type of method be utilized by social studies teachers or researchers.
- Literature reviews, in which the author(s) presented a summary of previous research and/or commentary on a topic or issue.
- Reaction papers, in which the author(s) reacted to critiques of their work that had appeared in an earlier issue of the journal.
- Validity or instrument development studies.
- Book reviews.

Of some 133 articles published in TRSE for this period, 87 (65 percent) fell into the above categories. Of some 73 articles published in the JSSR and another 33 published in the research section of SE for the same period, 18 (26 percent) and 16 (49 percent), respectively, fell into the above categories. These types of articles were not reviewed because they did not lend themselves to the kind of analysis we performed. We intend, therefore, no implication of the quality of these articles in any way by their omission.

A total of 118 articles in the three journals were reviewed. Table 1 gives a breakdown of the studies by type. The categories listed in the table were defined as follows:

Table 1
BREAKDOWN OF STUDIES BY TYPE

Type of Study	TRSE	JSSR	SE	Total
Pre-experiments	0 (0%)	4 (6%)	2 (11%)	6 (5%)
True experiments	7 (15%)	11 (17%)	5 (26%)	23 (18%)
Quasi-experiments	7 (15%)	9 (14%)	8 (42%)	24 (19%)
Correlational studies	9 (19%)	10 (16%)	0 (0%)	19 (15%)
Surveys	9 (19%)	23 (37%)	3 (16%)	35 (27%)
Interviews	6 (13%)	1 (2%)	1 (5%)	8 (6%)
Causal-comparisons	0 (0%)	3 (5%)	0 (0%)	3 (2%)
Ethnographies	9 (19%)	2 (3%)	0 (0%)	11 (8%)
Totals	n=47[a]	n=63[a]	n=19[a]	129[a]

[a]These totals exceed the actual number of studies reviewed because, in a few instances, two methodologies were used in the same study.

Figure 1. Categories used to evaluate social studies research

1. **Type of Research**
   - a. Experimental
     - 1) Pre
     - 2) True
     - 3) Quasi
   - b. Correlational
   - c. Survey
   - d. Interview
   - e. Causal-comparative
   - f. Ethnographic
2. **Justification**
   - a. No mention of justification
   - b. Explicit argument made with regard to worth of study
   - c. Worth of study is implied
   - d. Any ethical considerations overlooked?
3. **Clarity**
   - a. Focus clear?
   - b. Variables clear?
     - 1) Initially
     - 2) Eventually
     - 3) Never
   - c. Is treatment in intervention studies made explicit?
   - d. Is there a hypothesis?
     - 1) No
     - 2) Yes: Explicitly stated
     - 3) Yes: Clearly implied
4. **Are Key Terms Defined?**
   - a. No
   - b. Operationally
   - c. Constitutively
   - d. Clear in context of study
5. **Sample**
   - a. Type
     - 1) Random selection
     - 2) Representation based on argument
     - 3) Convenience
     - 4) Volunteer
     - 5) Can't tell
   - b. Was sample adequately described? (1 = high; 5 = low)
   - c. Size of sample (n)
6. **Internal Validity**
   - a. Possible alternative explanations for outcomes obtained
     - 1) History
     - 2) Maturation
     - 3) Mortality
     - 4) Selection bias/Subject characteristics
     - 5) Pretest effect
     - 6) Regression effect
     - 7) Instrumentation
     - 8) Hawthorne or John Henry effect
     - 9) Order effect
   - b. Threats discussed and clarified?
   - c. Was it clear that the treatment received an adequate trial? (in intervention studies)
   - d. Was length of time of treatment sufficient?
7. **Instrumentation**
   - a. Reliability
     - 1) Empirical check made?
     - 2) If yes, was reliability adequate for study?
   - b. Validity
     - 1) Empirical check made?
     - 2) If yes, type:
       - a) Content
       - b) Concurrent
       - c) Construct
8. **External Validity**
   - a. Discussion of population generalizability
     - 1) Appropriate
       - a) Explicit reference to defensible target population
       - b) Appropriate caution expressed
     - 2) Inappropriate
       - a) No mention of generalizability
       - b) Explicit reference to indefensible target population
   - b. Discussion of ecological generalizability
     - 1) Appropriate
       - a) Explicit reference to defensible settings (subject matter, materials, physical conditions, personnel, etc.)
       - b) Appropriate caution expressed
     - 2) Inappropriate
       - a) No mention of generalizability
       - b) Explicit reference to indefensible settings
9. **Were Results and Interpretations Kept Distinct?**
10. **Data Analysis**
    - a. Descriptive statistics?
      - 1) Correct technique?
      - 2) Correct interpretation?
    - b. Inferential statistics?
      - 1) Correct technique?
      - 2) Correct interpretation?
11. **Do Data Justify Conclusions?**
12. **Were Outcomes of Study Educationally Significant?**
13. **Relevance of Citations**

- **Pre-experiments.** We use this label to refer to any of the three types of "weak" research designs first described by Campbell and Stanley (1963): the one-shot case study, the one-group pretest-posttest design, and the static-group comparison design. As suggested by Stouffer almost four decades ago, studies employing such designs have such little control that they have almost no scientific value (Stouffer 1950).
- **True experiments.** Two or more groups of subjects receiving different treatments were compared in some way. Random assignment of subjects to treatment and control groups was assured. Administration of the treatment was controlled by the researcher.
- **Quasi-experiments.** Two or more groups of subjects were compared in some way. Random assignment of subjects to treatment and control groups did not occur. Administration of the treatment variable may or may not have been controlled by the researcher.
- **Correlational studies.** The scores of one group of subjects on two different measures were correlated. Such subsequent analyses as multiple regression or path analysis may have been performed.
- **Surveys.** A written questionnaire or test was administered, either by mail or in person, to one or more groups of subjects. No treatment was involved. The responses of the subjects were reported.
- **Interviews.** An interview schedule was prepared and administered orally (under the supervision of the researcher) to one or more groups of subjects. No treatment was involved. The subjects' responses to the questions were reported.
- **Causal-comparisons.** Two or more groups differing in known ways were compared on one or more variables. The intent was to explore possible causation between group membership and the other variable(s).
- **Ethnographies.** The daily activities of one or more individuals were studied in naturalistic settings. These activities, and manner of performing same, were described in detail. Case studies, involving only a single individual, were included in this category.

We acknowledge that this typology is imperfect; some studies, for example, involved more than one methodology. We decided to classify a study, therefore, according to the method or methods used to study the relationships or issues of interest. If more than one methodology was used, we counted both. We did not classify those studies that used analysis of covariance under "Correlational," however, since the use of correlation is an adjunct to the question of interest – the comparison of means. Furthermore, although ethnographic research may, and often does, incorporate interview procedures, we did not count this under "Interviews" since we believe this is generally understood.

Procedures

The instrument used for the analysis is shown in Figure 1. The categories listed therein were defined as follows:

1. **Type of research** – see discussion above.

2. **Justification of study** – the degree to which the worth of the study was explicitly argued for and/or defended. We also looked to see if there were any ethical considerations involved (i.e., whether there might be any physical or psychological harm to the subjects) and if so, whether the author(s) took such into account.

3. **Clarity** – the degree to which the study was clear. The concern here was with the focus of the study – its purpose and direction, and the degree to which (and when) the author(s) identified the variables they were investigating. We also looked to see if, in intervention studies, the exact nature of the treatment was made explicit and if so, when. Finally, any hypotheses that existed were identified, and the degree to which they were made explicit was assessed.

4. **Definitions** – the degree to which important terms in the study were defined, and how.

5. **Sample** – the type, size, and adequacy of description of the subjects involved in the study.

6. **Internal validity** – the number of plausible alternative explanations we identified for any reported outcomes and the extent to which these alternatives were identified and discussed by the author(s). We also considered whether it was clear that a treatment (in the intervention studies) actually occurred and when it did, whether the length of time of the treatment could be considered sufficient to produce the effect(s) intended.

7. **Instrumentation** – the degree to which any and all instruments used were demonstrably reliable and/or valid. We considered in particular whether the investigator(s) conducted any form of reliability and/or validity check of the instruments used and, if so, whether these checks were adequate for their purposes.

8. **External validity** – the extent to which the findings of the study were generalizable beyond the particular sample studied. Considerations here included both population and ecological generaliza-

bility, when and where the author(s) generalized appropriately (and, if so, to whom), when and where they did not, and when they could not, whether they explained why.

9. **Distinction between results and conclusions**—the extent to which the author(s) clearly differentiated between their findings (empirical data) and the conclusions they arrived at based on their findings (subjective opinion).

10. **Data analysis**—correct, and appropriate, use and interpretation of both descriptive and inferential statistics.

11. **Legitimacy of conclusions**—whether limitations raised crucial questions about the conclusions drawn.

12. **Educational significance of the study**—our judgment of the importance of the study in practical or theoretical, as opposed to statistical, terms.

13. **Relevance of citations**—the degree to which works cited in the articles were germane to the research being reported.

Each of us independently read and evaluated every study. We then met and compared our analyses. We do not report agreement of independent scoring because, although we had disagreements, the great majority were either clear oversights by one of us or easily resolved. It would have been desirable to compare our analysis with the findings of a second set of evaluators, but this was not feasible.

Results of Analysis

In the remainder of this section, we present the results of our analysis, using the categories from Figure 1 to organize our remarks. Both descriptive summaries of our findings and our interpretation of them are reported, along with examples of both good and bad practice.[1] We offer our observations on what these studies suggest about social studies research in Chapter 2.

Type of Research. The breakdown by type of research was shown earlier in Table 1. As can be seen, experimental and survey research predominate. This finding is in line with what other reviews have indicated (e.g., see Armento 1986; Stanley 1985). Of interest, however, is the preponderance of quasi-experiments in SE (some 42 percent of those reviewed); the rather large number of correlational studies in TRSE and the JSSR (almost 19 percent of the total number reviewed in TRSE and 16 percent in the JSSR); the equally large number of ethnographic studies in TRSE (19 percent of those reviewed); and the high percentage of questionnaire-type surveys in the JSSR (37 percent of those reviewed). One type of research methodology was particularly noticeable by its omission—*ex post facto* research.[2] We found not one example of this type of research published in any of the three journals during the period of this review.

It is worth noting that of the total number of articles published in TRSE during this period, 47 (35 percent) were arguments of one sort or another. This seems to be an unduly large proportion of the total number of articles published in this journal. The percentage of articles that were arguments was much lower in the other two journals.

Justification of Study. To what extent were these studies justified—that is, to what extent did the authors attempt to defend the worthwhileness of their research? A justification was considered to be any attempt by the authors either to argue explicitly why they thought their study was worth doing or clearly to imply its worth through their remarks.

The great majority of researchers did make an explicit argument for the worth of their research and did not simply take it for granted. Only five studies, in all three journals, did not contain some form of argument about the worth of the intended research. With regard to the ethics of these studies, in only one (out of 118) did we find cause for concern. This did not involve potential harm to the subjects, however, but rather what we considered to be inappropriate value judgments pertaining to another culture. The results in this category are shown in Table 2.

Clarity. The clarity of these studies received a mixed review (Table 3). We were pleasantly surprised to find that the focus—the overall intent—of *every* study was clear. We had no trouble whatsoever discovering what the authors intended to investigate. The clarity of the particular variables being investigated, however, was not always made clear. To be sure, in the great majority of studies in all three journals, the variables were made clear at the start. In seven of the studies in TRSE, however, it took us some time to be sure about the nature of the variables involved; in another eight, we never could discern what the variables were.

Of the eight studies in which the variables were unclear, five were ethnographies. Since one of the claims made for ethnographic research is the elucidation of meaningful variables, this falling seems rather serious. The authors of the remaining four ethnographies published in TRSE did succeed in making their variables clear, however. We recognize that one purpose of ethnographic research is sometimes said to be the presentation of ways in which differing groups give meaning to their existence—that is, their perceptions of the world. We did not, however, detect this intent in any of the ethnographies that we reviewed.

TABLE 2
RESULTS: JUSTIFICATION OF RESEARCH

Justification of Research	TRSE	JSSR	SE	Total
No mention of justification	2 (4%)	3 (6%)	0 (0%)	5 (4%)
Explicit argument made	35 (76%)	44 (80%)	17 (100%)	96 (82%)
Implicit argument found	9 (20%)	8 (14%)	0 (0%)	17 (14%)
Totals	46	55	17	118
Ethical concerns	0 (0%)	1 (2%)	0 (0%)	1 (1%)

TABLE 3
RESULTS: CLARITY

Clarity of Studies	TRSE	JSSR	SE	Total
Focus clear				
Yes	46 (100%)	52 (94%)	17 (100%)	115 (97%)
No	0 (0%)	1 (2%)	0 (0%)	1 (1%)
Questionable	0 (0%)	2 (4%)	0 (0%)	2 (2%)
Totals	46	55	17	118 (100%)
Variables				
Clear initially	31 (67%)	51 (93%)	17 (100%)	99 (84%)
Clear eventually	7 (15%)	3 (5%)	0 (0%)	10 (8%)
Never clear	8 (17%)	1 (2%)	0 (0%)	9 (8%)
Totals	46	55	17	118 (100%)
Treatment in intervention studies made explicit				
Yes	12 (26%)	17 (31%)	12 (71%)	41 (35%)
No	2 (4%)	5 (9%)	1 (6%)	8 (7%)
Not applicable	32 (70%)	33 (60%)	4 (23%)	69 (58%)
Totals	46	55	17	118 (100%)
Hypothesis present				
No	18 (39%)	24 (44%)	4 (24%)	46 (39%)
Yes, explicit	13 (28%)	11 (20%)	5 (29%)	29 (25%)
Yes, implied	15 (33%)	20 (36%)	8 (47%)	43 (36%)
Totals	46	55	17	118 (100%)

Of the studies published in the JSSR, the variables were clear in all but one (98 percent). The variables were clear in all of the articles published in the research section of SE (100 percent). Generally, too, in those studies involving an intervention of some sort, the treatment was made explicit, although there were a few in each journal in which we could not be sure as to what the treatment actually involved.

Twenty-eight of the 46 studies (61 percent) in TRSE, 31 of 55 (56 percent) of those in the JSSR, and 13 of 17 (76 percent) in SE were hypothesis-testing investigations. In over half of these in all three journals, however, the hypothesis was implied (e.g., in the rationale for the study) rather than being stated explicitly.

Definitions. Definition of key terms by the authors of these studies also drew a mixed review (Table 4). Almost 30 percent of the studies in TRSE and the JSSR lacked any definition of the terms involved; the figure is over 40 percent for SE. Interestingly, a disproportionate number (16 of 35) of these studies were either true or quasi-experiments. This was especially true in TRSE (7 of 13) and in SE (7 of 7!) It may be that since these studies tended to be on more traditional topics, using technical terms frequently found in the research literature, the authors assumed that these terms would be understood by the readership. This assumption may be questionable, however, and needs to be considered carefully.

Exactly 50 percent of the studies in TRSE (23 of 46) utilized either operational or constitutive definitions of terms (or both), compared to 40 percent for the JSSR and 18 percent for SE. The extent to which the meaning of the terms involved eventually became clear within the context of the study varied across journals, from 35 percent to 47 percent. Almost all of the TRSE studies having clear-in-context definitions occurred in the first half of the studies chronologically, whereas most of the studies that lacked definitions (10 of 13) occurred in the more recent 23 studies, allowing us to conclude that, overall, the adequacy of definitions decreased during this time period in this journal. This trend was not evident in the other two journals.

Sample. Only seven studies (out of a total of 118) had truly random samples (i.e., an initial random selection from a defined population). Three of these populations were so narrow as to be of dubious interest, however. They were (1) enrollees in teacher education at a particular university, (2) students from two high schools in the midwest, and (3) students from two high schools in the southeast. Three were surveys involving questionnaires, but the number of returnees totaled only 76 percent, 57 percent, and 80 percent, thus making the accepting sample no longer random. Finally, one involved a cluster sample with an "n" of only six classrooms, although these were randomly selected. The great majority were convenience samples, which, given the difficulties involved in doing research in the public schools, may (usually) be about the best one can expect. Table 5 shows the breakdown by type of sample.

The description of the sample often left a great deal to be desired (Table 6). Many times we were not clear about the characteristics of the subjects involved in a study.

TABLE 4
RESULTS: DEFINITIONS

Definitions	TRSE	JSSR	SE	Total
No definitions	13 (28%)	15 (27%)	7 (41%)	35 (30%)
Operational definitions	10 (22%)	6 (11%)	1 (6%)	17 (14%)
Constitutive definitions	13 (28%)	16 (29%)	2 (12%)	31 (26%)
Definitions clear in context	16 (35%)	25 (45%)	8 (47%)	49 (42%)
Totals	52[a]	62[a]	18[a]	132[a]

[a]Totals do not equal the actual number of studies reviewed, since several studies used both operational and constitutive definitions. The percentages represent percentage of the total of actual studies reviewed in which the particular type of definition (or absence of definitions) could be found.

TABLE 5
BREAKDOWN BY TYPE OF SAMPLE

Type of Sample	TRSE	JSSR	SE	Total
Total population	0 (0%)	1 (2%)	0 (0%)	1 (1%)
Random selection	2 (4%)	4 (7%)	1 (5%)	7 (6%)
Representation based on argument	6 (13%)	5 (8%)	2 (11%)	13 (11%)
Convenience	29 (62%)	42 (70%)	16 (84%)	87 (74%)
Volunteer	4 (9%)	3 (5%)	0 (0%)	7 (6%)
Can't tell	6 (13%)	5 (8%)	0 (0%)	11 (9%)
Total	47[a]	60[a]	19[a]	126[a]

[a]Eight studies used more than one type of sample. Percentages represent percentage of the total of actual studies reviewed in which the particular type of sample was used.

TABLE 6
RESULTS: ADEQUACY OF SAMPLE DESCRIPTION

Adequacy of Sample Demographics	TRSE	JSSR	SE	Total
Adequate sample demographics given	8 (17%)	1 (2%)	0 (0%)	9 (7%)
Some sample demographics given	29 (63%)	22 (40%)	6 (35%)	57 (48%)
No sample demographics given	9 (20%)	32 (58%)	11 (65%)	52 (45%)
Totals	46	55	17	118 (100%)

The adequacy of sample descriptions is an issue that is insufficiently discussed in the research literature. Is there any agreement that certain demographics, such as gender, age, socioeconomic status, or geographic area, for example, should always be reported? We know of no consensus on this question. Further, descriptive information must surely depend on the nature of the study. Perhaps authors should be required to report evidence that their sample is similar to a defined target population on variables they consider important. Perhaps it is unrealistic to expect satisfactory description. If so, another argument is raised in favor of replication; similar results obtained in several samples is an impressive argument for generalizability. Four studies did report some form of replication on the same topic – effects of teacher enthusiasm (three in TRSE and one in SE). Another four studies reported partial replications of a particular method of sequencing examples and non-examples in concept attainment. Interestingly, seven out of eight of these replicated studies involved a common investigator.

The lack of randomness in selecting samples and inadequate sample description raise serious questions about the generalizability of almost all the studies we reviewed; we shall discuss this point in more detail when we consider external validity.

The sample sizes in these studies varied tremendously, ranging from an n of one in an ethnographic study, to n's of 589 in an experimental study, 1800 in a correlational study, and 4150 in a questionnaire-type survey. The range of sample size by type of study is shown in Table 7.

Internal Validity. We were interested in how often alternative hypotheses could be suggested to explain positive findings. Accordingly, we examined each study to see the extent to which one or more threats to internal validity, originally identified by Campbell and Stanley (1963) and Cook and Campbell (1979), might have been present. Often, they were.

We acknowledge that this catalog of threats was originally developed to apply to experimental or group-comparison studies. As such, some of them make little sense when applied to correlational, questionnaire, interview, or ethnographic research (pretest, maturation, regression, and order effects, in particular). However, the remaining categories are useful with respect to all methodologies wherein a researcher is attempting to explore relationships and even (on occasion) when simple description is the aim. We believe the examples discussed below will document this point.

The most frequent threats were subject characteristics (other characteristics of the subjects may have accounted for the results), mortality (some of the subjects dropped out of one or more comparison groups in actual or probable unequal amounts), a Hawthorne or John Henry effect (the subjects in the experimental or control groups knew they were part of an experiment of some sort), and, especially in the ethnographic studies, a researcher effect (the researcher may have acted so as to bias the responses of the subjects in some way).[3] Furthermore, when these threats existed, the researchers oftentimes did not seem to be aware of them, or at least they failed to discuss their implications (this tended to improve somewhat in the more recent studies).

Table 8 shows the number of studies of each type in which we identified threats and (subsequently) the number where we judged them to be adequately discussed. Surprisingly, 14 of the 22 true experiments contained one or more threats. These included actual or probable inequality of groups despite random assignment (n = 7); lack of actual control over the treatment (n = 11); a possible Hawthorne or John Henry effect (n = 5); mortality (n = 4); and an instrumentation effect (n = 5). About half of the studies contained discussions of these threats. We were surprised that only one of the ethnographic reports acknowledged the possibility of a researcher effect, perhaps because it is thought to be an intrinsic limitation.

That we identified fewer threats, proportionately, for survey studies is not surprising, in that most of these studies attempted essentially to describe variables rather than to identify relationships. The other finding of possible importance is that the authors of studies in SE appeared to do a somewhat poorer job of discussing threats, possibly a reflection of the shorter length of these reports.

TABLE 7
RANGE OF SAMPLE SIZE BY TYPE OF STUDY

	TRSE		JSSR		SE	
Type of Study	**Range**	**(Med.)**	**Range**	**(Med.)**	**Range**	**(Med.)**
Pre-experiments	0		16-35	(31)	29	([a])
True experiments	42-589	(211)	24-282	(122)	18-360	(55)
Quasi-experiments	49-925	(200)	35-563	(74)	38-426	(164)
Correlational studies	33-1050	(498)	26-1800	(163)	0	
Surveys	25-554	(234)	16-2097	(93)	42-4150	([a])
Interviews	7-70	(27)	26	([a])	16	([a])
Causal-comparisons	0		120	([a])	0	
Ethnographies	1-138	(12)	3-26	(16)	0	

[a]Medians are not reported, since they have little meaning when there are only one or two studies.

TABLE 8
RESULTS: THREATS TO INTERNAL VALIDITY

A. Total Number of Threats to Internal Validity Identified in Each of Three Journals[a]

	TRSE	JSSR	SE	Total[b]
History	4	4	6	14 (12%)
Maturation	0	1	0	1 (1%)
Mortality	10	7	4	21 (18%)
Subject characteristics	15	31	8	54 (46%)
Pretest effect	2	6	1	9 (8%)
Regression effect	0	1	0	1 (1%)
Instrumentation	21	23	3	47 (40%)
Hawthorne/John Henry effect	7	7	10	24 (20%)
Order effect	0	1	1	2 (2%)

B. Types of Studies in Which Threats Were Identified and Discussed: TRSE

Type	No.[b]	Threats Identified[c]	Threats Discussed[d]
Pre-experiments	0 (0%)	0 (0%)	0 (0%)
True experiments	7 (15%)	3 (43%)	2 (29%)
Quasi-experiments	7 (15%)	7 (100%)	4 (57%)
Correlational studies	9 (19%)	5 (56%)	3 (33%)
Surveys	9 (19%)	3 (33%)	0 (0%)
Interviews	6 (13%)	4 (67%)	1 (17%)
Causal-comparisons	0 (0%)	0 (0%)	0 (0%)
Ethnographies	9 (19%)	9 (100%)	0 (0%)
	47[e]		

C. Types of Studies in Which Threats Were Identified and Discussed: JSSE

Type	No.[b]	Threats Identified[c]	Threats Discussed[d]
Pre-experiments	4 (6%)	3 (100%)	1 (33%)
True experiments	11 (17%)	8 (73%)	2 (18%)
Quasi-experiments	9 (14%)	9 (100%)	4 (44%)
Correlational studies	10 (16%)	10 (100%)	5 (50%)
Surveys	23 (37%)	11 (48%)	4 (17%)
Interviews	1 (2%)	1 (100%)	0 (0%)
Causal-comparisons	3 (5%)	3 (100%)	2 (67%)
Ethnographies	2 (3%)	2 (100%)	0 (0%)
	63[f]		

Table 8 (continued)

D. Types of Studies in Which Threats Were Identified and Discussed: SE

Type	No.[b]	Threats Identified[c]	Threats Discussed[d]
Pre-experiments	2 (11%)	1 (100%)	0 (0%)
True experiments	5 (26%)	4 (80%)	1 (20%)
Quasi-experiments	8 (42%)	8 (100%)	1 (13%)
Correlational studies	0 (0%)	0 (0%)	0 (0%)
Surveys	3 (16%)	0 (0%)	0 (0%)
Interviews	1 (5%)	0 (0%)	0 (0%)
Causal-comparisons	0 (0%)	0 (0%)	0 (0%)
Ethnographies	0 (0%)	0 (0%)	0 (0%)
	19[g]		

[a]Some studies contained several threats.

[b]Percentages represent the percentage of the total of actual studies reviewed within each category.

[c]The numbers and percentages here refer to studies in which threats were identified by us.

[d]The numbers and percentages here refer to a discussion by the authors of a study of the threats.

[e]One study used more than one methodology.

[f]Eight studies used two methodologies.

[g]One study used two methodologies.

We offer the following illustrations of how threats to internal validity may appear in other than comparison group studies. Whenever two or more instruments are used in a study, with both designed to investigate a particular relationship, an instrumentation threat may develop. There is sometimes a strong likelihood that at least some respondents will figure out the hypothesis and alter their responses accordingly, sometimes in ways making support for the hypothesis more likely. We viewed this as a problem in studies correlating (1) student self-concept, attitudes toward social studies, and perceptions of teacher and classroom, and (2) teacher attitudes toward teaching, self-concept, and acceptance of responsibility for student achievement.

An instrumentation effect may also occur due to the way instruments are administered and/or scored. In one study, for example, the report was such as to raise questions about the independence of scoring of the two instruments used to test the hypothesis; in another, the same administrator gave both tests to individual children, one following the other. In both studies, the instruments themselves were vulnerable to variations in administration and scoring.

The selection of subjects can create bias if the nature of the sample is atypical. This is closely related to the problem of generalizing, but is an additional problem in studies where it appears likely that the way subjects were obtained favors support for hypotheses. We judged this to be a problem in studies that (1) reported correlations between teacher responsibility for student achievement and various other attitudes in a sample of volunteers for a workshop in mastery learning, (2) reported relationships between student out-of-school experiences and attitude toward social studies in schools described as "good" in terms of environmental opportunities and quality of teachers, (3) reported correlations between a cloze reading test

and a test of text comprehension with a group of low socioeconomic level students whose teachers had low expectations for them, and (4) reported correlations between general concept attainment and understanding of social concepts with a group of primary children in a university lab school.

We also identified a possible subject selection threat in several survey studies, including (1) differences in male and female interest in social science disciplines using a sample from one area in the South, (2) opinions regarding effects of policies on research with human subjects, based on responses from a volunteer sample of "interested" faculty members, (3) teacher perceptions as to the nature of discipline problems in one school in a low income neighborhood, where discipline was considered a major problem, and (4) social action activities of social educators using a sample of volunteer respondents.

A positive sign with regard to internal validity was that it was generally quite clear in the intervention studies that the treatment was implemented as intended. We found only nine studies (out of a total of 50) in which this was not clear. Table 9 presents our impressions related to whether a treatment really did occur.

Whether the length of time of the treatment was sufficient to produce the intended effects proved to be a very difficult judgment to make, but we judged the time to be sufficient in only 60 percent of the studies (Table 10). Sizable differences among the journals appeared. In only four of the 14 intervention studies reported in TRSE did it seem that the prescribed treatment was clearly long enough to give the hypothesized effects an adequate chance to manifest themselves. This problem was less frequent in the other two journals, at least partly because the interventions themselves were often less "ambitious."

TABLE 9
RESULTS: TREATMENT

Was It Clear That a Treatment Occurred?	TRSE	JSSR	SE	Total
Yes	12 (26%)	17 (31%)	12 (71%)	41 (35%)
Questionable	2 (4%)	6 (11%)	1 (6%)	9 (8%)
Not applicable	32 (70%)	32 (58%)	4 (23%)	68 (57%)
Totals	46	55	17	118 (100%)

TABLE 10
RESULTS: LENGTH OF TREATMENT

Was Length of Time of Treatment Sufficient?	TRSE	JSSR	SE	Total
Yes	4 (9%)	19 (35%)	9 (53%)	32 (27%)
Questionable	9 (19%)	4 (7%)	4 (24%)	17 (14%)
Can't tell	1 (2%)	0 (0%)	0 (0%)	1 (1%)
Not applicable	32 (70%)	32 (58%)	4 (24%)	68 (58%)
Totals	46	55	17	118

Instrumentation. In this category, we were concerned with the extent to which researchers ascertained the reliability and validity of the instrument(s) they used. We looked to see if authors made some sort of reliability and/or validity check and, in the case of reliability, whether the reliability reported was adequate for the type of study being conducted. Those studies for which the answer to this query was "no" or "questionable" reported indexes below the rather lenient standard of .70. Here, as in other categories, results were not homogenous.

It is somewhat sobering to note that in all three journals, more than half of the studies did not make any reliability check whatsoever. This was the case with 25 of the 46 studies reviewed in TRSE; 29 of the 55 in the JSSR; and 10 of the 17 in SE. We judged reliability to be adequate in only 27 percent of these studies, including three studies reporting *only* scorer or observer agreement. We could find only four in which the researchers checked the stability of scores over time, probably a more important issue than internal consistency; only one of these reported the time interval involved.

Our findings with regard to validity were even more depressing. A startling 32 studies (out of 46) in TRSE; 46 studies (out of 55) in the JSSR; and 13 studies (out of 17) in SE made no attempt to check instrument validity! Of the 27 which did, only 12 presented evidence other than judgments. A more detailed breakdown on these data for all three journals is shown in Tables 11 and 12.

External Validity. External validity, of course, refers to the degree to which the results of a study are generalizable. This category was another in which the studies reviewed were distressingly deficient. Both population and ecological generalizability were considered in this category. Population generalizability refers to an explicit extension of the findings of the study to one or more target populations (i.e., other *subjects).* Ecological generalizability refers to an explicit reference to another *setting* of some sort (i.e., subject matter, materials, physical conditions, personnel, etc.).

In 22 instances in TRSE, the researchers generalized to indefensible target populations, although the authors did caution about generalizing inappropriately in another 13 studies. There was no mention of population generalizability in eight studies. Inappropriate generalizing occurred in 33 studies in the JSSR, and in 12 in SE. In both these journals, the frequency of entries in the "no mention of generalizability" category was considerably higher than those under "explicit reference to indefensible population." While this is clearly preferable, our experience indicates that to most readers, failure to discuss generalization leads to the erroneous inference that findings can be generalized without serious reservation.

TABLE 11
RESULTS: RELIABILITY OF INSTRUMENTS

Reliability	TRSE	JSSR	SE	Total
Empirical check made?				
No	25 (54%)	29 (53%)	10 (59%)	64 (54%)
Yes	21 (46%)	26 (47%)	7 (41%)	54 (46%)
Totals	46	55	17	118 (100%)
If yes, adequate for study?				
Yes	12 (26%)	16 (29%)	4 (24%)	32 (27%)
Questionable	4 (9%)	7 (13%)	1 (6%)	12 (10%)
No	4 (9%)	3 (5%)	2 (12%)	9 (8%)
Can't tell	1 (2%)	0 (0%)	0 (0%)	1 (1%)
Not applicable	25 (54%)	29 (53%)	10 (58%)	64 (54%)
Totals	46	55	17	118 (100%)

TABLE 12
RESULTS: VALIDITY OF INSTRUMENTS

Validity	TRSE	JSSR	SE	Total
Empirical check made?				
No	32 (70%)	46 (84%)	13 (76%)	91 (77%)
Yes	14 (30%)	9 (16%)	4 (24%)	27 (23%)
Totals	46	55	17	118 (100%)
If yes, type:				
Content (logical)	3 (21%)	0 (0%)	0 (0%)	3 (11%)
Judge-supported	2 (14%)	8 (89%)	4 (100%)	14 (52%)
Concurrent	5 (36%)	0 (0%)	0 (0%)	5 (19%)
Predictive	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Construct	3 (21%)	0 (0%)	0 (0%)	3 (11%)
Other (including factor analysis)	3 (21%)	1 (11%)	0 (0%)	4 (15%)
Totals	16[a]	9	4	29

[a]Two studies used two checks. Percentage represents percentage of studies in which instrument validity was checked.

There was no mention of ecological generalizability in 31 studies in TRSE, 44 in the JSSR, and 12 in SE, leading us to conclude that this, perhaps, is not something that these researchers generally considered. When they did, however, they were quite a bit more careful, with only six studies in TRSE, none in JSSR, and two in SE containing an explicit reference to an indefensible setting. We believe that the overall failure to discuss the ecological generalizability of a study, however, has the effect of suggesting that such generalizing is warranted. The breakdown in the three journals is shown in Table 13.

Distinction Between Results and Conclusions. Did the authors of these studies maintain a distinction between their findings (i.e., what they observed or obtained) and their interpretations (i.e., the conclusions they drew based on the nature of their findings)? Overwhelmingly, they did. Seventy-four percent of the studies in TRSE, 94 percent of those in the JSSR, and 82 percent of those in SE maintained a sharp distinction between results and interpretations. This is shown in Table 14.

The major exception was the ethnographic studies, which account for nine of the fifteen "no's." Although this is a widely known and, to some extent, unavoidable limitation of this type of study, we feel the authors of these studies could have done a much better job of making clear the basis for their interpretations. Failure to do so provides ammunition for those who allege that ethnographic research is little more than subjective impressionism.

Data Analysis. In almost all of the studies, the authors used some form of descriptive or inferential statistics. Did they use the correct procedure? Generally, yes! The five "no's" for descriptive statistics in TRSE and the two in the JSSR reflect our opinion that additional descriptive procedures (e.g., frequency of response) would have greatly clarified the findings. Three of the five "no's" in TRSE were ethnographies.

Were the interpretations of these researchers appropriate given the nature of their studies? Here, the answer generally is "yes" when descriptive statistics were involved, but overwhelmingly "no" when inferential statistics were reported. The major error was the inappropriate use of inferential

TABLE 13
RESULTS: EXTERNAL VALIDITY

	TRSE	JSSR	SE	Total
Discussion of Population Generalizability				
Appropriate:				
Explicit reference to defensible target population	2 (4%)	5 (9%)	2 (12%)	9 (8%)
Appropriate caution expressed	13 (28%)	16 (29%)	3 (17%)	32 (27%)
Inappropriate:				
No mention of generalizability	9 (20%)	23 (42%)	7 (41%)	39 (33%)
Explicit reference to indefensible target population	22 (48%)	11 (20%)	5 (30%)	38 (32%)
Totals	46	55	17	118
Discussion of Ecological Generalizability				
Appropriate:				
Explicit reference to defensible settings	2 (4%)	1 (2%)	0 (0%)	3 (2%)
Appropriate caution expressed	7 (15%)	8 (15%)	3 (18%)	18 (15%)
Inappropriate:				
No mention of generalizability	31 (67%)	44 (80%)	12 (71%)	87 (74%)
Explicit reference to indefensible settings	6 (13%)	0 (0%)	2 (12%)	8 (7%)
Not applicable	0 (0%)	2 (3%)	0 (0%)	2 (2%)
Totals	46	55	17	118

TABLE 14
RESULTS: DISTINCTION BETWEEN RESULTS AND CONCLUSIONS

Distinction Observed Between Results and Conclusions?	TRSE	JSSR	SE	Total
Yes	34 (74%)	51 (93%)	14 (82%)	99 (84%)
No	12 (26%)	3 (5%)	1 (6%)	16 (14%)
Questionable	0 (0%)	1 (2%)	2 (12%)	3 (2%)
Totals	46	55	17	118

procedures to test the significance of obtained results in studies where the obtained sample was not random. A significance test is appropriate only when a researcher is assured that he or she has a random sample, and this was literally the case in only four studies. (See the discussion under **Sample**.) In 13 other studies, the authors argued for representativeness and hence (by implication) for significance tests; we found only two of these persuasive. Some researchers advocate the reporting of significance tests as an indication of important differences but with appropriate qualifications. The reporting of effect sizes, however, we think would be more informative. Effect size was reported in only one study.

With regard to other forms of inferential misinterpretation, the author of one study in TRSE made much of the relative contribution of different variables to a multiple correlation, even after explicitly discussing the likelihood of chance fluctuations due to the small n (22). The authors of another, otherwise commendable, study committed the error of treating non-significant differences as though the null hypothesis were proven. In fact, the differences between the highest group and each of the two lowest groups were such as to yield effect sizes of approximately .4 to 1.0, depending on which standard deviation was used. This mistake also appeared in three other studies.

We found several studies in all three journals in which the authors apparently confused random assignment with random selection. Random assignment is a powerful, though imperfect, technique for equating groups. Further, it permits comparison of variance between groups with variance within groups. It does not, however, justify the calculation of significance tests, because generalization is a separate issue from both the equating of groups and assessing the magnitude of differences. When reporting a significant difference between two groups equated by random assignment, the question is: "To what population may this difference be generalized?" In the absence of random sampling, or of a persuasive argument for representativeness, and particularly in the case of convenience samples, which were used in virtually all of these studies, the answer must be: "No one knows!" Therefore, the information presumed in the finding of significance is, at best, only somewhat informative and, at worst, misleading unless carefully clarified by the authors, a practice glaringly absent from these reports, probably because it is virtually impossible to do.

In 28 studies, the interpretation of the descriptive statistics used was highly questionable. Nine of these were variations of correlation studies. Two of these combined scores of students with scores of their teachers in obtaining first-order correlations in multiple correlation studies, a highly suspect practice (particularly with a teacher n of eight in one study). In one of these two, it appears that data on teachers and students were simply combined; in the other, the best we can deduce is that the teacher's scores were assigned to each of his or her students. The author of the latter study also concluded that the obtained results provided limited support for the position that teachers should be encouraged to focus their instruction around objectives. This conclusion was based on the finding that teacher use of objectives contributed one percent (*one* percent!) to the predicted variance of student achievement (whereas the CAT and pretest combined contributed 42 percent)! In both studies, the unnecessary complexity of analysis and reported data virtually preclude the reader from determining what the findings really were.

In six other studies, too much was made of correlations below .40. While a case may sometimes be made for the importance of correlations of this magnitude in testing theory or in unusual practical applications (e.g., prediction with a very small

selection ratio), one can hardly pay serious attention to correlations of this size when the variables are "historical understanding" and "information processing capacity" (r=.14); "economic knowledge" and "attitude toward the American economic system" (r=.28); "positive interracial contact" and "satisfaction with university life" among black females (r=.22); and IQ vs. close-mindedness and self-esteem (r=-.24; .29), even though statistically significant due to large n's. Another study states that "some modest school effects were found for political interest, political alienation, and anti-Vietnam war attitudes." The multiple correlations based on five school variables *plus* IQ and socioeconomic level were, respectively, r=.39; r=.16; and r=.41—modest indeed, especially since the particular schooling variables were weighted differently for each attitude. Granted, the low reliability of instruments used may limit the degree of correlation possible; this is another reason for reporting reliabilities. In the absence of such data, however, one cannot assume that correlations would be higher with more reliable instruments. One study illustrates this point in reverse. The author dismisses a correlation of .53 because it is reduced to .32 when a subgroup of restricted range is analyzed. This, when the respective reliabilities of his two instruments are maximally .64 and .96![4]

We found 12 studies using the group comparison model which contained highly questionable interpretations. In one quasi-experimental study, the authors concluded, on the basis of non-significant t tests (n=49 in each group), that there was a "lack of major effects on the attitude of MACOS students," while admitting that the MACOS group became slightly more tolerant of repugnant activities than did the non-MACOS group. Examination of the change in total test score means, however, shows that the MACOS group showed a change of -2.94 compared to -.48 for the comparison group. Estimation of the standard deviation of change scores for the comparison group suggests an effect size of .6 to .7, an impressive difference even though not statistically significant. The authors of another, otherwise well-done, experimental study concluded that one of four teaching strategies was the most useful and devoted considerable space to discussing why this might be so. This, despite the finding that this was the *poorest* of the four methods for one of their four interaction subgroups (female poor readers), while another method was appreciably better. The authors of this study also committed the error of assuming that non-significant differences on pretests is tantamount to groups being equal. Regressed gain scores should have been used, since pretests were given expressly to check on the efficacy of random assignment in equating groups.

One of two hypotheses tested in a quasi-experimental study was that regular value analysis discussions would increase students' social trust, social integration, political confidence, and political interest, as compared to reading-only and control groups. Under the results section of the study, the authors concluded that "there is some evidence to support the hypothesis." They go on to state that while the value analysis group did score significantly better statistically than the reading-only group, the difference between the two groups was minimal. In addition, the control group scored significantly higher than did the reading-only group on two of the measures. They then concluded that the results offered only modest and mixed support for the hypothesis. In actuality, the adjusted means for the value analysis and control groups were very similar. The only meaningful finding is the lower scores for the reading-only group. The authors provide plausible interpretations as to why this group may have scored lower while the control group scored high, but such *ex post facto* speculation cannot obviate the finding that there was *no* support in the data for the hypothesis. In another study, a low correlation between pre and post scores on an attitude scale (single group) was interpreted as indicating true change after eight weeks of summer school, rather than the more probable low reliability of the scale.

In most of the surveys where we questioned the interpretation, the reason was lack of supporting data. In one case, however, the author simple ignored data that was presented. By combining the categories "agree" and "slightly agree," the interpretation of differences in attitude toward different social studies traditions was, in fact, obscured.

For the most part, the errors described above appear to support the opinion, increasingly voiced (e.g., see Carver 1978; Shaver and Norton 1980), that inferential statistics play too important a role in current research efforts. Not only are they, with rare exception, mathematically or logically indefensible, but they also can obscure the the real findings of a study. Perhaps it is time for the profession to consider using descriptive statistics more meaningfully, rather than continuing to foster the use of elegant but inappropriate inference tests.

The breakdown with regard to the analysis of data in these studies is shown in Table 15.

Legitimacy of Conclusions. Were the conclusions reached by the authors of these studies justified? This, perhaps, is the most important question addressed in this review. In attempting to answer it, we decided to focus on the extent to which the conclusions drawn by the authors seem

TABLE 15
RESULTS: DATA ANALYSIS

Descriptive Statistics	TRSE	JSSR	SE	Total
Use correct?				
Yes	34 (74%)	51 (92%)	17 (100%)	102 (86%)
No	5 (11%)	2 (4%)	0 (0%)	7 (6%)
Questionable	1 (2%)	2 (4%)	0 (0%)	3 (3%)
N/A[a]	6 (13%)	0 (0%)	0 (0%)	6 (5%)
Totals	46	55	17	118
Interpretation correct?				
Yes	26 (57%)	38 (69%)	14 (82%)	78 (66%)
No	9 (20%)	7 (13%)	0 (0%)	16 (14%)
Questionable	0 (0%)	9 (16%)	3 (18%)	12 (10%)
N/A[b]	11 (23%)	1 (2%)	0 (0%)	12 (10%)
Totals	46	55	17	118
Inferential Statistics				
Technique Correct?				
Yes	28 (61%)	35 (63%)	13 (76%)	76 (64%)
No	1 (2%)	0 (0%)	0 (0%)	1 (1%)
Questionable	0 (0%)	2 (4%)	0 (0%)	2 (2%)
Can't tell	0 (0%)	1 (2%)	0 (0%)	1 (1%)
N/A[a]	17 (37%)	17 (31%)	4 (24%)	38 (32%)
Totals	46	55	17	118
Interpretation correct?				
Yes	3 (7%)	1 (2%)	1 (6%)	5 (4%)
No[c]	26 (56%)	37 (67%)	11 (65%)	74 (63%)
Questionable	0 (0%)	0 (0%)	1 (6%)	1 (1%)
N/A[b]	17 (37%)	17 (31%)	4 (24%)	38 (32%)
Totals	46	55	17	118

[a]N/A indicates that statistics were not reported nor considered necessary.

[b]N/A indicates that statistics were not reported. In some cases we think they should have been.

[c]A rating of "no" indicates at the very least no mention of violation of the underlying assumption of random sampling.

defensible within the confines of the study itself. We deliberately excluded the important issue of generalizability, which would have resulted in a much more negative evaluation (see Table 13). The main factors influencing our judgment were: (1) adequacy of instrumentation, (2) severity of threats to the internal validity of the study, and (3) adequacy of the interpretation of data. (In addition to the weaknesses discussed previously, we found an all-too-common tendency to make cause-effect statements in much stronger terms than were justified.) In our judgment, the conclusions reached by the authors were clearly justified in only 20 (44 percent) of the studies published in TRSE, 27 (49 percent) of those in JSSR, and nine (53 percent) of those in SE (Table 16).

Educational Significance of Studies. Researchers often talk about the statistical significance of their findings, but they just as often fail to talk about the significance of their results in any larger sense. Why are the results of a study important, and to whom? What is the practical significance of a study's results? Why do they *matter* (or do they)? We asked ourselves these questions as we read these studies and attempted to weigh them in this light. Would the results of any of these studies make a difference to teachers and other professionals? In our judgment, many of them would not. We give our impressions in Table 17. The phrase "can't tell" indicates we were so confused by the study as to be unable to judge its significance. Although we almost always agreed, we acknowledge that we cannot clearly articulate the basis for this judgment.

Relevance of Citations. Table 18 indicates our judgment of the relevance of citations for the topic of a study. In our judgment, the references in some studies had little direct relevance to the study involved.

Notes

1. We do not cite specific studies critiqued because we have no desire to engage in destructive criticism or to embarrass authors. We will be happy, however, to provide citations to interested readers who wish to assess the accuracy of our specifics.

2. By *ex post facto* research, we mean any study in which an investigator seeks an explanation for findings that have already occurred. Suppose, for example, that an administrator in a large, urban high school notices that the end-of-year test scores for students in a particular social studies teacher's classes are markedly higher than those of the students of other teachers, and have been for several years. She wonders why, and decides to compare several variables of the two groups – characteristics of the students, materials used, teaching style, etc. – in an attempt to gain insight into why this is the case. The differential results, however, have already occurred, and the administrator is seeking an explanation for these results after the fact.

3. We include researcher effects under instrumentation.

4. Maximum $r_{1,2} = \sqrt{r_{1,1}\, r_{2,2}} = \sqrt{.64 \times .96} = .78$

TABLE 16
RESULTS: LEGITIMACY OF CONCLUSIONS

Were the Conclusions of the Study Legitimate?	TRSE	JSSR	SE	Total
Yes	20 (43%)	27 (49%)	9 (53%)	56 (47%)
No	13 (28%)	13 (24%)	1 (6%)	27 (23%)
Questionable	3 (7%)	14 (25%)	7 (41%)	24 (20%)
Can't tell	10 (22%)	1 (2%)	0 (0%)	11 (10%)
Totals	46	55	17	118

TABLE 17
RESULTS: EDUCATIONAL SIGNIFICANCE OF THE STUDIES

Were the Outcomes Educationally Significant?	TRSE	JSSR	SE	Total
Yes	22 (48%)	15 (27%)	7 (42%)	44 (37%)
Questionable	12 (26%)	13 (24%)	1 (5%)	26 (22%)
No	10 (22%)	27 (49%)	9 (53%)	46 (39%)
Can't tell	2 (4%)	0 (0%)	0 (0%)	2 (2%)
Totals	46	55	17	118

TABLE 18
RESULTS: RELEVANCE OF CITATIONS

Relevance (1 = high; 5 = low)	TRSE	JSSR	SE	Total
1	17 (37%)	10 (18%)	5 (29%)	32 (27%)
2	10 (22%)	22 (40%)	1 (6%)	33 (27%)
3	17 (37%)	19 (35%)	7 (41%)	43 (36%)
4	2 (4%)	3 (5%)	4 (24%)	9 (9%)
5	0 (0%)	1 (2%)	0 (0%)	1 (1%)
Totals	46	55	17	118

CHAPTER 2
SOME SUMMARY OBSERVATIONS ABOUT SOCIAL STUDIES RESEARCH

What does our analysis of these studies reveal? In general, progress over time appears to be slow. Much still can be done, it appears, to improve the quality of social studies research. We offer the following observations.

Methodology

Experimental and survey research methodologies predominate. Of the 118 studies we reviewed, 47 (40 percent) were either true or quasi-experiments, and 43 (36%) were either questionnaire- or interview-type surveys, for a total of 90 (76%). Recent reviews of research (e.g., Armento 1986; Stanley 1985) document that these types of research continue to dominate our field.[1] Other forms of research, such as historical inquiries and ethnographic studies, are much less commonly found, both in doctoral dissertations and in our research journals, although they do occur. Some research methodologies, such as causal-comparative investigations, are truly rare. In our review, we found only three (2 percent) studies that were causal-comparative investigations.

We think this is too narrow a vision of research to dominate the field. The term *research* means any sort of "careful, systematic, patient study and investigation in some field of knowledge, undertaken to discover or establish facts and principles" (*Webster's New World Dictionary* 1984). Many methodologies fit this definition. Additional models that could (and should, we think) be utilized by social studies educators more frequently include case studies; content analyses; intensive, in-depth interviews (particularly when used to illuminate student comprehension); historical inquiries; correlational studies; structured observations; participant observations; causal-comparative investigations; ethnographic studies; and cross-cultural comparisons. Out of 239 articles published in the three journals (of which we reviewed 118), only 20 were content analyses; 19 were correlational studies; 11 were ethnographies; eight were historical inquiries; three were structured observations; and three were case studies. There was one each of participant observations and cross-cultural studies.

We think that all of these research methodologies have value, since each constitutes a different way of inquiring into the realities that exist within social studies classrooms and the minds and emotions of social studies students, teachers, and other professionals. While all of these methodologies (as well as experiments and surveys) have various limitations (and thus can be well or poorly executed), their wider use would help to provide some additional, and different, perspectives about important questions in social studies education.[2] It is encouraging to note that, while they still remain relatively few compared to the more common forms of experimental or survey research, more studies using some of these alternative methodologies are being reported in the social studies research literature (Armento 1986; Stanley 1985).

Many research questions in the social studies can be studied through experiments or surveys, but they also might well be investigated by other methodologies. Indeed, some of the other methodologies that we have mentioned are often better suited to providing the information desired. We believe that research in social studies education should ask a variety of questions, move in a variety of directions, encompass a variety of methodologies, and use a variety of tools. Different research orientations, perspectives, and goals should not only be allowed, but encouraged.

Focus/Clarity

In general, the authors made clear the focus and variables of their studies. Definitions were somewhat better than expected, though lacking in some 30 percent of the studies. These researchers sought primarily to understand more clearly or in more detail various aspects of the field. The great majority did not try to point up inaccuracies, distortions, ideological bias, etc., but rather to understand more fully the outcomes of particular methods/techniques, the characteristics of students, and the characteristics and opinions of social studies professionals. Although this comment is in line with what other reviewers have observed, a growing amount of research of the former type is being reported (e.g., see Anyon 1978; Giroux and Penna 1979; Popkewitz, Tabachnick, and Wehlage, 1981; Romanish 1983; Saltonstall 1979). Although we found few empirical studies of a critical nature in the three journals we reviewed, we did find many arguments and position pieces (e.g., Cherryholmes 1982; Common 1982; Cornbleth 1985; Egan 1980; Giroux and Penna 1979; Gordon 1985; Hahn and Blankenship 1983; Holmes 1982; Hurst 1980; Romanish 1983; Stanley 1981; Wasburn 1986).

Sample

Only seven (6 percent) of the 118 studies reviewed attempted to use truly random samples, as compared with a total of 15 percent found by Shaver and Norton (1980). Sample descriptions often left much to be desired.

Replication

Unfortunately for the build-up of a knowledge base, we found only eight studies (6 percent) that were replications of other work (four on each of two topics). This continues to be a major failing of social studies research. The social studies research community has not made a systematic effort to build a cumulative base of knowledge about many of the important questions of interest to the profession. Doctoral students continue, in the main, to do isolated studies, often unaware that similar or related work is being done by their counterparts elsewhere (Hepburn and Dahler 1985). Few doctoral, and fewer master's, studies are expanded or developed further once they are completed. As Shaver has remarked, there is "a failure in many instances...to relate a piece of research to previous studies in any sort of programmatic way. The consequences are, on the one hand, the repetition of unproductive prior research and, on the other, a disconnectedness of studies on similar topics. Both are counterproductive to knowledge building" (Nelson and Shaver 1985, p. 410).

Internal Validity

The internal validity of many studies, unfortunately, was suspect. Threats that appeared in a large number of studies included a subject effect, in which characteristics of the subjects may have accounted for the results; an instrumentation effect, in which the data collection procedure may have acted to bias the results; a Hawthorne or John Henry effect, in which some of the subjects may have known they were part of an experiment of some sort; and mortality, where some of the subjects may have dropped out of the comparison groups in unequal amounts. A positive sign with regard to internal validity was that, in the intervention studies, it was clear that the treatment actually did occur.

Reliability and Validity of Instruments

Reliability and validity checks on instruments were not performed in a large majority of studies. Out of the 118 studies reviewed, 64 contained no reliability checks whatsoever; in *91* of these studies, the researchers made no attempt to check validity! These researchers appeared either to ignore these issues or to accept unquestioningly evidence from prior data collection. In some cases, such evidence did seem appropriate to the study at hand, but in many cases, unfortunately, it did not. The absence of reliability and validity checks continues to be a major failing in much social studies research.

External Validity

The external validity of these studies also proved deficient. In almost three-quarters of the studies in all three journals, the authors either explicitly or implicitly generalized to indefensible target populations; in 74 percent, they made no mention of ecological generalizability, thereby implying that it should be taken for granted. Although these authors generally used the correct statistics in analyzing their data, they often interpreted their findings incorrectly, leading us to conclude that many in the profession appear to lack adequate understanding of statistical interpretation.

Theory

Very few of the authors of these studies tried to connect their work to some underlying theory. While the usefulness of theory in guiding and organizing research can hardly be questioned, we doubt whether the diversity of our field can be encompassed in any one theory. Most likely, at least at present, we shall have to settle for theories that address subtopics, such as the Merrill-Tennyson theory of sequencing examples in concept attainment, which provided focus for one of the two replicated topics in our review.

Investigators

In virtually all the studies we reviewed, as well as in virtually all the articles *published* in the three journals (i.e., including the instrument studies, the position pieces, the content analyses, the historical studies, and any others that we did not review), the authors were college professors (primarily) or other social studies professionals (supervisors, state social studies specialists, etc.). Classroom teachers were noticeable by their absence. We found only two studies in which a classroom teacher was one of the researchers. Although we recognize the severe workloads under which most social studies teachers labor, we lament their absence from these reports. Accordingly, we shall offer some suggestions as to the kinds of research investigations classroom teachers could conduct in Chapter 5.

Data Sources

Where did these researchers get their data? In almost all of the studies we reviewed, the data collected by the researchers appeared to come from

one of four main sources: (1) the performance of students in social studies classes, (2) the opinions of students and/or teachers in schools, (3) the viewpoints of social studies supervisors, social studies methods professors, or other social studies professionals, (4) various documents, such as courses of study, curriculum guides, state frameworks, etc.

Another approach to the problem of basic data gathering is the development of centralized data bases that social studies researchers and others might use and to which they could contribute. The details of both existing and potential data bases are beyond the scope of this monograph, but three that should be mentioned are:

- National Assessment of Educational Progress (NAEP)
  Educational Testing Service
  Princeton, NJ 08540
- The College Board
  45 Columbus Avenue
  New York, NY 10023-6917
- High School and Beyond
  Center for Statistics
  Department of Education
  555 New Jersey Avenue, NW
  Washington, DC 20208-1310

Topics Investigated

What about content? What topics did these researchers study? How significant were these topics? Almost all of these studies focused on relatively narrow or (in our judgment) unimportant relationships, rather than on important concepts and ideas, or important issues. This observation is in line with what many other reviewers have noted (Ehman and Hahn 1981; Metcalf 1963; Shaver 1979b; Shaver and Larkins 1973; Wiley 1977; Stanley 1985). A positive comment is that in the majority of these studies, the authors did attempt to justify their research. We did a limited content analysis of the topics investigated by the authors of these studies. Our findings are shown in Table 19.

Such analyses, of course, obscure the specific questions addressed, but they do give a picture of overall activity. Within these categories, the only specific topics addressed in more than two studies were concept acquisition and development (12 studies), student political attitudes and opinions (five studies), student opinions on social studies content (five studies), effects of teacher enthusiasm (four studies), student knowledge of economics (four studies), and teacher evaluation of the use of objectives (three studies). Noteworthy by their absence were any studies that looked specifically at the learning of gifted students in social studies, that compared the social studies learnings of different (i.e., ethnic, cultural, socioeconomic, etc.) subgroups of students or that analyzed existing data bases of the type mentioned above. Also missing were investigations of social studies in different types of school settings (e.g., urban vs. rural, public vs. private, etc.), in specialized (e.g., magnet, classics-oriented, comprehensive, etc.) schools, or in other lands. Virtually all of the studies we reviewed investigated aspects of social studies in the United States. We found only two articles (of 239 published) in all three journals that described aspects of social studies education in another country.

The variety of topics covered is not surprising in a field that is by definition and tradition as diverse in its subject matter as social studies and as influenced by community expectations and values (Berlak and Berlak 1981). As several critics have shown (Barr, Barth, and Shermis 1977; Morrissett and Haas 1982; Newmann 1986; Stanley 1985), there is continuing disagreement among social studies theorists and curriculum developers as to what should be emphasized in our field. It is customary under such conditions to call for renewed attempts to unify the field or at least parts of it under some theoretical structure. We applaud such efforts, but we do not believe researchers can, or will, await such developments.

As an alternative, we suggest that leaders in the field attempt to identify and even prioritize important categories, topics, and/or questions for researchers to investigate. Ehman and Hahn (1981) proposed several categories in Part II of the 1981 National Society for the Study of Education Yearbook (see pp. 60-78). Nelson and Shaver (1985) proposed a list of questions in their chapter in the recent review of social studies research (see pp. 408-410). We suggest another below. We believe that the attention of researchers, above all else, should be directed toward finding out what students know and how to help them learn. This is hardly a new idea, but one worth repeating. Our preferences in this matter are strongly influenced by our work with the late Hilda Taba. We believe that both the content and process objectives she advocated provide a sound basis for focusing research. Whether one agrees with her curriculum approach, we believe the following questions are worth considering:

1. What is the present level and range of student knowledge and/or commitment at different ages/grade levels with respect to major objectives in social studies, namely:

   a. Identified concepts and ideas?

   b. Identified cognitive skills?

   c. Identified values?

TABLE 19
A CONTENT ANALYSIS OF RESEARCH TOPICS IN TRSE, JSSR, AND SE

Topic	TRSE	JSSR	SE	Total
Characteristics of teachers or supervisors (and their effects)	4	6	1	11
Attitudes or opinions of social studies educators	5	10	0	15
Characteristics of student teachers	2	5	0	7
Characteristics of students	15	8	3	26
Teaching methods	16	15	12	43
Pre- or inservice training	0	8	0	8
Program requirements	3	3	0	6
Dissemination of innovations	1	0	1	2
Totals	46	55	17	118

2. What is the present level and range of teacher capabilities with respect to:

a. Their own knowledge and attitudes toward these objectives?

b. Their competence and/or aptitude for teaching these objectives, especially with students of differing abilities, age, ethnicity, and gender?

3. What are the attitudes of parents, school personnel, and school boards toward these objectives?

4. What methods are effective in increasing understanding and support for these objectives among the groups named in question #3?

5. How do the objectives mentioned in question #1 correspond to developmental patterns in general cognitive abilities, interest, and attitudes among students?

6. How do students of differing cultural, ethnic, and socioeconomic backgrounds vary with regard to these objectives?

7. What general and specific teaching methods are effective in fostering these objectives with different types of students and in different types of schools and communities?

8. How can these methods accommodate important differences in student readiness? If necessary, how can important readiness variables be assessed?

9. How can teachers best be helped to implement these methods? To what extent must they develop their own methods or adaptations?

10. What factors, both within and without the school, hinder and help the education of students with respect to these objectives?

11. To what extent does competence in these objectives generalize to other subject matter and to daily-life activities?

Clearly, this is just a beginning. However, we believe some such systematic approach to our field would permit researchers to focus their efforts in ways that would contribute to a more integrated body of knowledge with relevance to important issues of policy and practice. We also believe that failure to study these issues is one reason why the "new" social studies projects of the 1960s had less impact than was anticipated.

In conclusion, our analysis supports recent criticism of educational research in general as being deficient in both application and discussion of principles of good research with respect to sampling, internal validity, instrumentation, and data analysis. We also concur that both topics and methodology are too narrow. On the positive side, we were pleasantly surprised at the general quality of justifications, clarity of focus and terminology, documentation of treatment implementation, adequacy of treatment time, and distinction between results and interpretations.

Notes

1. These also are the methodologies most commonly found in social studies doctoral dissertations. Based on a review of the abstracts of some 394 doctoral dissertations written between 1977 and 1983, Hepburn and Dahler found that descriptive studies comprised 45 percent, or 177 of the total. Experimental research comprised 27 percent, or 105 of the total. Thus, the two together totaled 282, or 72 percent of the total (Hepburn and Dahler 1985, pp. 77-78).

2. We also recommend the analysis of existing data bases, which we believe is an area of research that has been largely ignored to date by the social studies research community (see page 25 for a list of a few of these data bases.)

CHAPTER 3
SUGGESTIONS FOR IMPROVING THE QUALITY OF SOCIAL STUDIES RESEARCH

In light of the foregoing analysis and observations, we wish to offer some suggestions we think could improve the quality of social studies research. The remarks in this section are directed to three distinct groups of social studies educators: (1) professors who direct master's theses or doctoral dissertations in social studies education, but who do not teach courses in educational research, (2) graduate students in social studies education who intend to do research, and (3) classroom teachers, curriculum directors, and administrators who have an interest in research.

Before we offer our suggestions, however, we wish to make a distinction between the terms *social education* and *social studies*, for the remarks that follow have mainly to do with social studies research. The distinction to which we subscribe is the one offered by Nelson and Shaver: *social education* is "a term inclusive of the broad concerns of social knowledge, social relations, social development, and social improvement, which are among the goals of social studies, but go beyond schooling practices in their intentions, activities, and research implications," whereas *social studies* identifies "the schooling part of social education" (Nelson and Shaver 1985, p. 401). Most of our suggestions apply primarily to research in schools. Although some may apply to studies that take place outside of schools, such studies are not the focus of these remarks.

What follows, then, are some ideas about how to improve the quality of social studies research. Many of these ideas are suggested by the weaknesses we noticed in our review in Chapter 1. Since experiments and surveys remain the most commonly conducted types of research (72 percent of the total in the studies we reviewed), more of our suggestions focus on these methodologies than others. Space limitations prevent an extensive discussion of other methodologies, but we offer a few ideas that frequently seem to be ignored in practice. Since most social studies educators are not trained in historical or ethnographic research, these methodologies in particular seem to be logical candidates for further study (e.g., see Agar 1986; Barzun and Graff 1977; Bogdan and Biklen 1982; Carr 1967; Dobbert 1982; Gottschalk 1969; and Spindler 1982).

All of the ideas we present are relatively easy to implement. Very few are new; most have been identified by one or more other observers. Nevertheless, we believe that they bear repeating. Experience with our own graduate students suggests that even those students who have had two or three courses in research continue to make rather fundamental mistakes. Furthermore, as our analysis in Chapter 2 revealed, many of these ideas continue to be ignored in practice.

Improving Experimental Research

1. De-emphasize random sampling. Obtaining a truly random sample is almost an impossibility in school-based research, given today's organizational and scheduling constraints. When and where possible, of course, random sampling is to be encouraged. An alternative strategy, however, is to concentrate on describing relevant demographics of one's sample (e.g., ages, gender, ethnicity, IQ scores) in enough detail so that other researchers (and other interested professionals) get a fuller picture of exactly who was involved in the study. We believe the profession might profitably attempt to develop guidelines as to the kind of description that ought to be provided.

Oftentimes, even in intact classes, random assignment of students to treatment and control groups can be implemented. It should be recognized, however, that this technique is really only effective with large groups (we recommend at least 50 subjects per treatment group). When smaller groups must be used, or when random assignment is not feasible, much more attention should be given to matching (mechanically or statistically[1]) groups on potentially related variables, as well as on the outcome (dependent) variable(s).

2. Increase the chances of the treatment's having an effect. In essence, this suggestion involves *intensifying* the treatment the experimental group receives. There are three possibilities here:

a. Be clear that there *is* a treatment. Sometimes treatments are so vaguely defined or described that exactly what happened to students in the experimental group is not clear. Operational definitions of the independent variable(s) can help clarify the nature of the treatment.

b. Lengthen the *time* of the treatment. Oftentimes, the length of time that students are exposed to a treatment is so short that its possible effect(s) may not be discerned (Wallen and Fraenkel 1988). Eisner (1983) found that the median experimental treatment time per subject in the studies that he reviewed in 1978 was only 45 minutes! One can take slight encouragement from the fact that a

review some five years later showed an increase in the median experimental treatment time to one hour and 15 minutes per subject (Eisner 1983, p. 14). Although the treatment time was considerably longer in the studies we reviewed, it was still of concern in over one-third of the total.

c. Check (through the use of observers, audio- or videotaping, subject reactions, etc.) to make sure that the treatment *really occurs* and that it occurs as intended.

3. Concentrate on description and explanation more than prediction. Given the difficulty in obtaining random samples in most school settings, the generalizability of most social studies research will be severely limited. This suggests the value of placing more emphasis on description and explanation and less on prediction. Vividly described details of interventions (or in non-intervention studies, of settings) can help others in similar situations assess the applicability of particular results to their situation. As mentioned above, the nature of the treatment should be clearly and fully described. Exactly what happened? How? When? Where? Under what conditions?

4. Use more than one instrument to measure the dependent variable. In the great majority of social studies research, the researchers use only one measuring device to obtain data concerning the outcome of interest. This unnecessarily limits the amount of information gathered concerning the possible effects of the independent variable(s). Use of a second instrument also permits a check on concurrent validity. In our review of 118 studies that used instruments, only 12 (10 percent) used more than one measuring device to obtain data on the dependent variable(s).

5. Pay more attention to alternative explanations of findings due to "mortality" and "Hawthorne effect" threats. We found a sizable number of studies in which these were concerns (16 and 20, respectively). If subjects are lost to a study, researchers should attempt to determine whether the proportion was about the same for all treatments and whether the causes were likely to favor certain treatment groups. If, for example, the experimental treatment is a difficult one for students, and hence those "lost" are those having the most difficulty (they may change groups or just absent themselves from treatment or testing), the data on those remaining in that group would not reflect the lower performance of the absentees.

A possible Hawthorne effect exists whenever one group receives any sort of special attention. This threat is hard to control in studies involving a major curriculum modification, since provision for special treatment of comparison groups is often not feasible (or is artificial). Despite the difficulties presented by these two threats, they should receive more attention than appears currently to be the case.

6. Study more than one dependent variable. Rarely do social studies researchers look at more than one dependent variable when studying the effects of a particular treatment. Once again, this unduly restricts the amount of information that might, with only a little extra effort, be obtained. It also weakens understanding of the possible effects of an independent variable. Theory or experience usually suggests that a treatment will affect more than one outcome variable. Further, unintended or unanticipated outcomes should be studied to the extent feasible. It is not very difficult, for example, to measure the attitudes of students in studies where achievement is the dependent variable (e.g., see Smith 1980). We are *not* suggesting that additional variables be included merely for the sake of addition. A clear and defensible rationale is always required.[2]

7. Incorporate additional independent variables into your design. Many times the effect(s) of a treatment may be predictably revealed in one or more subgroups, yet not appear in the total group of which the subgroups are a part. Analyzing a treatment group in terms of gender or ethnic components, for example, may reveal otherwise unrecognized effects. Factorial designs that enable a researcher to study several independent and dependent variables in a single study are almost never employed in social studies research.

8. Discuss the magnitude of any effects observed. Social studies researchers commonly report their findings in terms of significance levels, using inferential statistics, but the notion of statistical significance is intimately related to sample size. Given a large enough sample, almost any result will be statistically significant. Whether a finding is significant only tells us the likelihood of an effect occurring by chance; it does not allow us to compare effects across studies of similar phenomena. As many observers have suggested, the calculation of an *effect size* is helpful in this regard (Borg and Gall 1983; Nelson and Shaver 1985; VanSickle 1983). Similarly, the reporting of the percent of variance accounted for – Eta^2 – provides another indication of magnitude.

9. Be less concerned about statistical significance and think more about educational significance (despite the difficulty of assessing the latter). The significance of a study continues, for most social studies (and other) researchers, to mean statistical significance. Because the results of a study are statistically significant (were not due to chance), however, does not mean that they are significant in any larger sense. The import of a study –

how it matters in the larger scheme of things, to students, to teachers, to the profession as a whole—is rarely discussed. Researchers should watch for *noticeable* effects whether they are statistically significant or not.

In particular, the emotional reactions of students should be assessed if at all possible. How strongly did they react to a particular treatment or experience? Why do they say they react in this way? Do different groups react differently? When students react strongly (either positively or negatively) to an intervention or an experience, further investigation is probably warranted. Of the 118 studies we reviewed, only five (4 percent) assessed student reactions to social studies subject matter; only two (1.6 percent) assessed student attitudes toward some aspect of social studies instruction!

10. Assess the durability of an effect. Delayed posttests are virtually never given to see whether the perceived effects of an independent variable remain over any length of time or change in any way (Leming 1985). The durability of the effects of independent variables in social studies research remains largely unknown.

11. Make better use of descriptive statistics. Whether we are correct in believing that one of the causes is overemphasis on inferential statistics, it is clear that many of the studies we reviewed inappropriately used and/or interpreted basic descriptive indices. We agree with Kerlinger (1986) that excessive reliance on computer packages may be further contributing to this problem. We encourage researchers to stay closer to their data and pay greater attention to such simple indices as medians (in addition to means), as well as to frequency polygons and scatterplots—both of which can be easily obtained through computer analysis. We recommend that much more thought be given to both the magnitude and the pattern of group differences found and their implications—which may be quite different for questions that are primarily theoretical as compared to practical.

12. Give more attention to the interpretation of results. Most of the studies we reviewed did a good job of keeping results and interpretations distinct. Most often, however, the larger meaning of results was inadequately discussed. Too often, authors discussed unwarranted direct applications (population generalizability).

We would recommend more discussion of the implications of results in the context of both practice and theory. For example, a study finding that understanding of social studies ideas is correlated with level of general concept development in young children implies that (a) teachers may need to assess level of concept development, and (b) developmental theories with regard to concepts apply to social studies content. Such discussion would help others decide whether the replication needed to generalize is worth the effort.

Improving Survey Research

1. Trial-test all questionnaires or interview schedules. Of the 43 survey studies we reviewed, only two indicated that the questionnaire or interview schedule used was checked beforehand. Pilot testing with a small group similar to the group to whom the questionnaire or interview schedule is to be administered can help reveal lack of clarity, bias, and/or ambiguity in questions before it is too late to change them.

2. Check the validity and reliability of the questionnaire or interview schedule being used. Many studies reporting survey results do not indicate if, or how, the validity and reliability of the survey instrument were checked. Like any measuring instrument, a questionnaire or interview schedule needs to be checked for reliability and validity to insure that data obtained is related to what the researcher is trying to assess. Out of the 118 studies we reviewed, only 21 (46 percent) in TRSE, 26 (47 percent) in the JSSR, and seven (41 percent) in SE made some attempt to check instrument reliability, while a startling 32 (70 percent) in TRSE, 46 (84 percent) in the JSSR, and 13 (76 percent) in SE made *no* attempt to check validity! Content validity, at least, can be assessed through the use of independent judges who rate the questions to be asked in terms of whether they measure the variables the researcher has in mind. The researcher can then revise any to which the judges object.

Many investigators appear to think that validity is unimportant when factual questions are asked. They need to remember that it is not the fact itself that is of concern, but the *way in which the factual information is obtained*. This certainly can lead to invalid interpretation. It is often difficult, but not impossible, to ask for the same factual information in more than one way, as Kinsey and his associates demonstrated forty years ago (Kinsey, Pomeroy, and Dartin 1948).

3. Think about the length of the questionnaire or interview schedule. It should be neither too long nor too short. The proper length, of course, is a matter of judgment, but researchers need to consider whether their instruments are sufficiently long to provide them with enough information concerning what they are looking for, yet not so long that respondents become tired, bored, or careless. The length of a survey instrument may seem too obvious a point to mention, but almost everyone has neglected to respond to a survey at least once because the length of the questionnaire discouraged us from doing so.

4. Check for sampling bias. How representative is the accepting sample (those who actually respond to the questions) of the specific group being surveyed? This depends, of course, on the percentage of responses returned. When a substantial percentage of responses is not received (we think more than 20 percent), representing the findings as indicative of the invited sample may be misleading. (This happened in many of the survey studies we reviewed.) A possible check on this is to interview a small sample of nonresponding subjects to see how, or if, their views differ markedly from those of the respondents. A second (or even a third) administration of the questionnaire can also help increase the percentage of responses returned. Showing that respondents are similar to invitees with respect to at least some demographic variables permits additional confidence in generalizing findings.

5. Check respondent knowledge about the subject before or during administration of the questionnaire or interview schedule. This is to make sure that respondents actually possess some knowledge concerning what they are to be questioned about. Otherwise, the researcher cannot be sure that their replies represent what the respondents actually know about the issue(s) being surveyed.

6. Try to make sure that you and your respondents speak the same language. Several years of experience in helping students design questionnaires have shown us that this cannot automatically be assumed. Sometimes a particular term can mean the exact opposite of what the researcher intends. Babbie (1986, p. 230) described an example in which the word "very" in the colloquial language of Appalachia apparently was closer to what people in other parts of the country mean by "fairly" or even "poorly." Thus, when residents of the area responded "very well" to an inquiry about their health, they actually meant that they were just getting along. The best "solution" to this problem is a prior tryout that includes questions (preferably in interview form) specifically directed toward the meaning of terms.

7. Train all individuals who will administer an interview schedule to ensure that they are able to administer it correctly. Such training helps ensure that the data obtained will be both reliable and valid. Training should include a trial run to check on the manner of administration. Use of videotapes to provide feedback can be very helpful.

8. Try to make sure that both researcher and respondents are operating from the same frame of reference—that is, respondents must be clear about what the researcher expects regarding the questions being asked. This guards against differential expectations leading to erroneous interpretations by the researcher. For example, if a researcher were to ask, "What do you think about what goes on in your history class?", one student might talk about the kinds of activities used by the teacher; another might comment on the homework assignments; yet another might talk about the teacher's way of questioning students. Others, unsure of what the questioner wants, might not respond at all. A less ambiguous question might be: "What do you think of the way your teacher conducts class discussions?" The important point here is that the researcher must make clear to respondents exactly what he or she wants them to respond to or comment about.

9. Don't use an observation form with too many categories. Researchers must take care that their observational measuring instruments (e.g., tally sheets, flow charts) are neither too long nor too short. Overly long observation instruments require too much of observers, while overly short ones produce only a partial analysis of what is observed. The difficulty involved in using an overly complicated tally sheet has been the downfall of many a graduate student.

10. Check on the interrater agreement of independent observers to ensure a high degree of reliability (we would argue for at least .90). Reliability should be reported, using internal consistency indices where appropriate. Stability over time should also be checked.

11. Be sure to take a random or systematic sampling of whatever is being observed. Observing just the beginning of a class, for example, can mislead researchers. Many reports of observations in social studies classrooms do not make clear exactly when, or during what period of time, the observations took place. Typically, a sizable number of observation periods (eight or more) is necessary to achieve adequate reliability.

Improving Correlational or Causal-Comparative Research

1. Be careful not to imply that correlation indicates causation. Although the fact that correlation does *not* mean causation is one of the most frequently mentioned caveats in research courses and research texts (e.g., Borg and Gall 1983; Kerlinger 1986; Vockell 1983; Wallen 1974; Wiersma 1987), many studies still imply, on the basis of a significant correlation, that a cause-and-effect relationship exists.

2. Don't confuse statistical significance with educational (or practical) significance. This error is similar to that found so often in experimental studies. Interpretation of the magnitude of a correlation coefficient continues to be one of the most

misunderstood aspects of research in social studies education. Correlational coefficients ranging from .20 to .35 show only a *slight* relationship between variables, even though they may be statistically significant. A correlation of .20, for example, indicates that only four percent of the variance in the two variables that have been correlated is common to both. Such correlations have almost no value in any practical sense. A correlation of at least .50 must be obtained before any crude predictions can be made concerning groups (although they are usually of little help in making individual predictions). Even then such predictions are frequently in error (since they indicate only a 25 percent common variance). It is only when a correlation of .65 or higher is obtained that individual predictions that are reasonably accurate for most purposes can be made. Correlations over .85 indicate a close relationship between the variables correlated and are useful in predicting both group and individual performance, but correlations this high are rarely obtained in social studies research (Borg and Gall 1983).[3]

3. Analyze as many relevant subgroups within the total sample being studied as possible. Many times, important relationships may be obscured when correlations are computed just for the total sample, rather than for certain subgroups within it as well. Sizable correlation coefficients may be found when subgroups (e.g., males and females) are examined. In analyzing subgroups, researchers should also examine the variability within each, since this affects the magnitude of the correlation.

Improving Ethnographic Research

1. Reflect on your own subjectivity. Ethnographers have wrestled for years with the criticism that a researcher's biases can influence his or her descriptions. All research can be affected by personal bias. The task for all of us is to *limit* our bias. One way to do this in ethnographic research is to take into account one's biases by describing, in detail, one's *thoughts* about what one is observing; in effect, to write memos to oneself about what one is thinking (Bogdan and Biklen 1982).

2. Do your best to "blend into the woodwork." The subjects of a study often attempt to create a false impression of themselves, especially during the early stages. Teachers might not yell at any students, for example, or be especially patient. Students may be unusually cooperative. Principals may disrupt their normal routines. Accordingly, the researcher needs to act in such a way that the activities and conversations that occur in the researcher's presence are no different from those occurring in his/her absence. A thorough understanding of the research setting is therefore crucial. Certain data may not ring true. Some data, in fact, may need to be discounted once it is interpreted in context (Deutscher 1973).

3. Be a conversational rather than a formal questioner. This idea is related to the suggestion above. A conversational form of interchange with subjects is more likely to engender natural, non-staged responses than is formal administration of an interview schedule or questionnaire.

4. Take care that you are not unduly influenced by the most talkative subjects. Oftentimes, a researcher talks with certain students a disproportionate amount of time compared to other students for the simple reason that they are the most willing to talk. This can result in misleading impressions and interpretations. You need not talk with all subjects for the same amount of time, but you should not rely exclusively on only a small number of subjects whose ideas may be somewhat atypical. Less talkative subjects should not be given up on too quickly.

5. When appropriate, share your feelings about experiences you observe with your subjects. A researcher's feelings can help him or her establish rapport with subjects and gain insight into their feelings. Bogdan and Biklen described an instance in which an observer was overwhelmed with a feeling that things were out of control in a junior high school cafeteria she was visiting for the first time. When she mentioned her feelings in the teachers' room, several teachers began to discuss their feelings during their first few weeks on cafeteria duty. Discussing her feelings enabled the observer to gain insight into the feelings of the teachers in this school that she otherwise might never have obtained (Bogdan and Biklen 1982, p. 132).

6. When observing, practice describing rather than interpreting. Anthropologists work very hard at training themselves to avoid placing their own inferences into their basic data. No competent anthropologist, for example, would write in his or her field notes: "Ms. Jones punished Robert," which is clearly an inference, but rather something like "Ms. Jones told Robert to be still," or "Ms. Jones sent Robert to the office." Unfortunately, many applications of this methodology in education, including all of the ethnographies we reviewed, appear to be vulnerable to this criticism.

7. Make a major effort to check information from more than one source (e.g., observations with interviews, interviews with different informants). While this is a basic technique for validating all information, it is especially important in ethnography, since so much interpretation by the researcher is required.

Some Ideas for Improving Research in General

1. Make greater use of volunteers as subjects in methods studies. It is standard advice that use of volunteers is a serious threat to the generalizability of a study and hence should be avoided. This is true, but it is important to note that a negative result (in intervention studies) when volunteers are the subjects is a strong statement concerning the effectiveness of the treatment. If a treatment does not work with volunteers (whom we would assume would be more motivated than most), this is a pretty good indication it will not be effective with most other subjects. Perhaps this should be the first step in studying innovative methods.

2. Consider the context within which a study takes place. Much experimental and quasi-experimental research, for example, involves only one classroom, at most a very few, in which a treatment is applied under atypical conditions. Hence the applicability of the results to what most social studies classroom teachers do on an ongoing basis is often hard to see (this may be one of the reasons why most classroom teachers pay little attention to social studies research). Furthermore, little attention is usually paid to the nature of the school environment within which most teachers work, and whether it would be possible for teachers to manipulate students in ways similar to manipulation in research studies. Although we did not specifically evaluate studies on this issue, our overall impression is that virtually none addressed the issue of context.

3. Indicate how the research relates to previous studies of the question at issue. Oftentimes there is no tie-in made to other, related work, nor any indication of what other researchers have found with regard to the same, or similar, questions. Attempting to relate one's own research efforts to the work of others is another contribution that social studies researchers could make relatively easily to the building of a cumulative knowledge base in the field. The variation shown in our assessments indicates our judgment that the studies we reviewed differed a great deal in this regard.

4. Formulate and state a hypothesis when appropriate. Many social studies researchers undertake their investigations without formulating and testing a prediction of some sort. Some critics would argue that the generation of hypotheses before a study begins limits the researcher's observations, in that he or she may overlook or ignore data not related to the hypothesis. The value of formulating a hypothesis, however, is threefold: (a) it forces us to think more deeply about what we want to investigate and often clarifies what outcome(s) we are looking for, (b) it stimulates us to begin thinking about how we can test our theories, and (c) it encourages the development of a body of knowledge. Many studies designed to investigate the same hypothesis but containing different moderator variables might contribute to the building of the knowledge base that the profession so badly needs, yet at present does not have. Of the 118 studies we reviewed, only 29 (25 percent) contained an explicitly stated hypothesis; another 43 (36 percent) contained an implied hypothesis. Forty-six (39 percent) did not attempt to investigate a hypothesis.

5. Be sure to define key terms clearly. The lack of clearly defined terms is one of the most common findings in the literature. In much social studies research, the reader is unsure as to what the researcher means by many of the terms he or she uses. Terms like *active learner, critical thinking, values development, citizenship education,* and others are frequently not defined. Thirty percent (35 out of 118) of the studies we reviewed lacked any definition whatsoever of the terms involved.

It would be helpful to define all key terms operationally—that is, to specify observable characteristics, behaviors, or conditions (along with how they can be measured). For example, defining *motivation* as a desire to learn is not very clear. A clearer definition would be: "any statements or actions an individual makes or takes which, in the judgment of at least two teachers or counselors, indicates the individual's desire to learn."

6. Remember that instrument reliability is crucial. Unless instruments are "sufficiently" reliable (a complex matter that can be reduced to the rule of thumb that the coefficient of reliability should exceed .70), you are probably wasting your time. Checking internal consistency is usually a simple matter. While other types of reliability do require more elaborate data collection, they should be seriously considered.

7. Pay more attention to the possibility of a researcher effect. Researchers can influence study outcomes by systematically (though unintentionally) favoring certain treatment groups in either treatment application, data collection or both. We found very little attention given to this issue by either ethnographers or more traditional researchers.

While it is true that standardization of procedures reduces this problem, a better guarantee of impartiality is ignorance, at least on the part of data collectors, who generally do not need to know the hypotheses or purposes of a study. In the studies we reviewed, the researchers appear to have been the data collectors in virtually all cases. While this may be legitimate and even necessary, the possibility of bias on the part of the researcher should at least be discussed.

8. Use more than one statistical tool to analyze findings. Here again, a little extra preparation and effort can pay dividends. Many, if not most, researchers use only one statistical procedure when they analyze their data. Most usually, means and standard deviations are computed. Frequently, additional statistics can be computed and presented; these include, as appropriate, percentages, medians, ranges, correlation coefficients, and effect sizes. These statistics can provide additional information as to how various groups compare (e.g., see Powell and Powell 1984).

9. Finally, do more to replicate previous research. Almost all research in social studies education is done in isolation. With rare exception (e.g., see Larkins and McKinney 1982), the replication of previous work under somewhat different settings, with different subjects or modified treatments, is simply not done. As Shaver has suggested, the systematic replication of research findings would not only help "to establish their reliability and generalizability" but also past research efforts could be used "as a basis for designing studies to correct methodological errors and build on past findings" (Nelson and Shaver 1985, p. 411). We find it hard to understand why the use of master's theses to replicate significant studies, a common practice in the physical sciences, has never caught on in the behavioral sciences in general, and in social studies research in particular.

Further, we recommend that more researchers *cross-validate* their research by checking their findings with the findings of others who used different methods. Thus, a researcher who found through interviews that teachers *said* they asked certain kinds of questions in class could check to see if this finding is consistent with the findings of another study using direct observation.

While this list is not intended to be exhaustive, we believe it highlights many of the more obvious weaknesses we noticed in our review. In order to discuss some of these suggestions further, we analyze a single study in some detail in Chapter 4.

Notes

1. We recommend calculation of regressed gain scores rather than (or in addition to) use of the very similar, but non-identical analysis of covariance because regressed gain scores provide additional descriptive information (the adjusted gain score for each student) as well as means and standard deviations.

2. We recommend that social studies researchers consider the sophisticated and potentially powerful techniques of confirmatory factor analysis and covariance structural analysis, which are combined in LISREL (Linear Structural Relations), a system incorporating computer analysis. These types of analyses permit elegant and satisfying clarification of some questions, but they do require considerable mathematical, statistical, and computer sophistication. They also require a degree of theoretical clarity that is currently lacking in our field. We would caution further that such techniques make many of our recommendations all the more important to consider.

3. We think the coefficient of determination (r^2) should also be reported. In addition, the reporting of beta weights permits evaluation of the magnitude of relationship analogous to the use of effect sizes with regard to means.

CHAPTER 4
A DETAILED ANALYSIS OF A SAMPLE STUDY

In an effort to point up some of the criticisms and observations made earlier in this monograph, in this chapter we dissect a single study, using the same instrument we used to analyze the various studies in TRSE, the JSSR, and SE. We discuss both the strengths and weaknesses of this study as a way of reinforcing some of the ideas we have presented for improving social studies research.

The study we analyzed was chosen for several reasons:

- Though the study itself was conducted quite some time ago, the focus of the study is current. Critical thinking has recently reemerged as a priority in the social studies curriculum. In addition, the teaching method used in the study remains a major approach supported by critical thinking advocates.
- The methodology used by the researchers is typical of current research efforts.
- The use of a study in which one of us was the lead author permits us to be more critical than we might otherwise choose to be.

The Study

The study is reproduced in its entirety below.

THE OUTCOMES OF CURRICULUM MODIFICATIONS DESIGNED TO FOSTER CRITICAL THINKING*

Norman E. Wallen, Vernon F. Haubrich, and Ian E. Reid, University of Utah**

CRITICAL THINKING appears to be a universally accepted objective of education though we are frequently unclear as to what we mean by it and to what extent we wish to live with its consequences. As has been pointed out elsewhere (5), various definitions of critical thinking seem to encompass some or all of the following features:

1. Use of scientific methods, including emphasis on evidence and the nature of hypotheses.

2. The tendency to be inquisitive, critical, and analytical with respect to issues, personal behavior, etc. A derivative of this attribute is lack of susceptibility to propaganda.

3. Use of correct principles of logic.

The emphasis is on the development of that elusive philosophical idea, the rational man.

With respect to methods of fostering critical thinking, two major approaches have been advocated. The first is "progressive education." Critical thinking is presumed to be but one of the objectives which are fostered by a greater degree of self-determination, flexibility of curriculum, and freedom of behavior. The results of the Eight Year Study provided some support for this position. Further support of an indirect type is provided by studies which indicate that questioning and critical behaviors are less likely to occur in rigid, highly formalized situations wherein deviation is punished (2).

The second approach emphasizes the tools rather than the attitude of critical thinking while recognizing the importance of a milieu conducive to the use of the tools. Thus, emphasis is placed on acquainting students with the principles of logic and experimentation and with their use. It is this approach toward which this study was directed.

---

*From the *Journal of Educational Research* (July-August 1963), pp. 529-534, Reprinted by permission of the authors.

**At the time of the study.

Method

The basic design of the study was as follows:

It involved seven teachers of U.S. History (eleventh grade) in three Salt Lake City high schools who introduced the curriculum modifications and an additional two who served as controls. During the first year, one class (selected at random) taught by each of the nine teachers was tested in the fall and again in the spring to establish the amount of gain to be expected over a year's time under the present curriculum. The tests used were the Cooperative U.S. History Test, the Watson-Glaser Test of Critical Thinking, and the I.D.S. Critical Thinking Test. During the summer of 1960, the experimental teachers attended a one week workshop on the University of Utah campus under the direction of Dr. Haubrich, during which time they received training in the curriculum procedures and materials presently available as well as experience in the development of new materials. During the following academic year, two of their classes were again tested in the fall and spring as were those of the control teachers. During this year, the staff members worked with the teachers in the utilization and development of materials. The resulting data permitted comparisons of gains made from year to year under the same teacher and from teacher to teacher within a given year.

The statistical analysis used was analysis of covariance, which permits comparison of end-of-year scores—adjusted for beginning-of-year scores under the different treatments. Thus (in effect) the mean gain achieved by the experimental teachers during the first year—regular curriculum—is compared with the mean gain achieved under the modified curriculum. Further, the mean gain achieved by the experimental teachers using the modified curriculum is compared with the mean gain achieved by the control teachers during the same year.

Curriculum Modifications

The overall plan of curriculum modification called for the teaching of a unit in "critical thinking" followed throughout the year by application to the content of the course as rather broadly defined. As an example, the students were encouraged to examine their textbook, their newspapers, and their teachers for examples of fallacious logic. This approach has been extensively developed in the Illinois Curriculum Program under the direction of B. Othanel Smith and his associates. In a comprehensive application of the plan in Illinois, a total of 36 teachers and approximately 1,500 high school students in English, geometry, science, and social studies classes participated. As of this writing, only a preliminary report has been published (5). It appears that the study was carefully conducted and that the students experiencing the experimental method showed greater gain on measures of critical thinking than the control group without showing impairment in mastery of course content.

Thus, the present study is, to a large extent, a replication of the Illinois study to determine whether similar results are obtained—a procedure woefully lacking in educational research. In addition the present study contains some methological improvements, notably the use of a "base line" for gauging change which is based on the same teachers who institute the curriculum changes.

For convenience, the curricular practices may be divided into (1) materials presented during the unit on critical thinking, and (2) materials used throughout the remainder of the year.

1. **Unit on critical thinking.** This unit required approximately three weeks for all teachers and was conducted—at the teachers' convenience—sometime during the second or third month of school.

The sequence of presentation varied from teacher to teacher but included the following topics and in this general order:

a. Definitions—abstract and concrete

b. Logical fallacies—*post hoc* fallacy, etc.

c. Deductive principles
   Syllogisms
   If—then statements
   Validity and truth

d. Inductive principles
   The nature of evidence
   Analysis of arguments including recognition of implicit assumptions
   Reliability of sources

In addition to their notes and experiences during the workshop, the teachers were provided with copies of *Applied Logic* by Little, Wilson, and Moore and copies of *Guide to Clear Thinking* developed by the Illinois Curriculum Program. Also, it was intended that each student be provided with or have access to *A Guide to Logical Thinking* by Shanner. In one school, however, a misunderstanding resulted in these booklets not being available to all students.

As can be seen from the topics listed above, the intent was to present to these students many of the more salient developments in the areas of logic, semantics, and philosophy of science, but in a fashion which they would comprehend.

2. **Application.** Throughout the remainder of the year, the teachers attempted to utilize the ideas and skills taught during the unit whenever feasible. To this end many of the exercises developed by the Illinois group were used. Also, the teachers showed considerable ingenuity and expenditure of effort in materials which they developed. Some of the flavor of the materials may be conveyed by the following illustrative exercises.

a. A statement on page 77 of the text states: "The Articles of Confederation granted considerable power to a Congress of the United States." Is this definition, explanation, or opinion? What criteria are provided?

b. Analyze the argument for unfair advantages of big business on page 368 of the text. Are there irrelevancies? Fallacies? Do the reasons justify the conclusion?

c. Is there a fallacy in the following argument? Life under a strong central government in Great Britain was tyrannical. We must not allow a strong central government to develop in this country.

Tests Used to Evaluate Outcomes

The measuring devices used to assess the outcomes of the program included the Watson-Glaser Critical Thinking Appraisal, the I.D.S. Critical Thinking Test, both constructed to assess skills in critical thinking, and the Cooperative U.S. History Test, which was used to assess change in the more typical content of the course.

1. **Watson-Glaser.** This test was originally published in 1942 and was revised in 1956. It contains five sub-tests: inference, assumptions, deductions, interpretation, and arguments. It has been used in numerous studies and is quite adequate in terms of technical considerations such as reliability, norms, etc. Ennis (3) has, however, questioned its validity on the grounds that some items are questionable and that it gives too high a score to the "chronic doubter."

2. **I.D.S. Test.** This test was developed in 1957 by Ennis, in part as an attempt to overcome his objections to the Watson-Glaser. As such the items are, on logical grounds, superior. Preliminary data suggest that it is adequate from a technical standpoint.

3. **Cooperative U.S. History Test.** This test is considered to be one of the best standardized tests of the typical content of American History courses. It contains items designed to test knowledge of historical facts; understanding of cause-and-effect relationships, trends and developments; and ability to recognize chronological relationships, interpret historical maps, and locate historical information with emphasis on political and diplomatic history. It is somewhat weak in the area of contemporary affairs.

Results

Results of the analysis of covariance comparing students of the experimental teachers for the two years are shown in Table 1. Table 2 shows the analysis of covariance comparing experimental and control classes for the second year only. Table 3 shows the means of the various groups as well as some additional data pertaining to the I.D.S. Test. Tables 4 and 5 show mean values for the Watson-Glaser and Cooperative U.S. History Test, respectively. These data support the following interpretation:

TABLE 1
ANALYSIS OF COVARIANCE – EXPERIMENTAL TEACHERS ONLY*

Source of Variance	Σx^2	Σxy	Σy^2	d.f.	Adj. Σy^2	M.S.	F	P
I.D.S. Test								
Between years (curricula)	4	27	189	1	159	159	8.83	< .01
Between teachers	423	347	428	6	168	28	1.56	
Interaction	187	229	489	6	287	48	2.67	< .05
Residual	8865	5066	10245	406	7350	18		
Total	9479	5669	11351	419				
Watson-Glaser								
Between years (curricula)	44	58	77	1	15	15	.32	
Between teachers	888	732	654	5	59	12	.26	
Interaction	1219	1374	1732	5	373	75	1.59	
Residual	28042	20298	31108	348	16414	47		
Total	30193	22463	33570	359				
Cooperative U.S. History Test								
Between years (curricula)	904	-104	12	1	899	899	18.65	< .001
Between teachers	1530	1235	1200	6	216	36	.75	
Interaction	634	725	1083	6	291	48	1.00	
Residual	23988	21601	39027	406	19575	48		
Total	27056	23457	41322	419				

*With the exception of the F column, decimals have been omitted to simplify the tables.

Cases were deleted at random so as to obtain samples of 20 each for each teacher for year 1 and 40 for each teacher for year 2. This procedure necessitated dropping the classes of one teacher from the Watson-Glaser analysis, since only 12 students took both test and re-test during year 1.

TABLE 2
ANALYSIS OF COVARIANCE – EXPERIMENTAL VS. CONTROL TEACHERS – YEAR 2 ONLY

Source of Variance	Σx^2	Σxy	Σy^2	d.f.	Adj. Σy^2	M.S.	F	P
I.D.S. Test								
Between groups	124	216	375	1	154	154	7.22	< .01
Residual	8203	5066	11352	386	8222	21		
Total	8327	5282	11727	387				
Watson-Glaser								
Between groups	254	206	167	1	24	24	.47	
Residual	12262	13789	35229	391	19790	51		
Total	12516	13965	35396	392	19814			
Cooperative U.S. History Test								
Between groups	226	314	435	1	49	49	1.09	
Residual	22198	20397	36224	386	17481	45		
Total	22424	20711	36659	387	17530			

TABLE 3
MEANS OF EXPERIMENTAL AND CONTROL GROUPS IN THE PRESENT STUDY AND OF OTHER COMPARISON GROUPS ON THE I.D.S. TEST

	N	Mean Fall	Mean Spring	Gain
Experimental Teachers - Regular Curriculum - Year 1	140	8.8	10.2	1.4
Experimental Teachers - Modified Curriculum - Year 2	280	9.1	11.8	2.7
Control Teachers - Regular Curriculum - Year 1	36	6.8	8.4	1.6
Control Teachers - Regular Curriculum - Year 2	53	7.5	9.0	1.5
Normative Data - High School Juniors*			9.0	
Normative Data - High School Seniors*			9.6	
College Educational Psychology Students*			12.3	
High School Students in Courses Emphasizing Critical Thinking*			12.1	

*Ennis, R.H. "Interim Report: The Development of the I.D.S. Critical Thinking Test."

TABLE 4
MEANS OF EXPERIMENTAL AND CONTROL GROUPS ON THE WATSON-GLASER TEST

	N	Mean Fall	Mean Spring	Gain
Experimental Teachers - Regular Curriculum - Year 1	120	62.3	64.9	2.6
Experimental Teachers - Modified Curriculum - Year 2	240	61.6	64.0	2.4
Control Teachers - Regular Curriculum - Year 1	30	56.8	60.0	3.2
Control Teachers - Regular Curriculum - Year 2	53	59.6	62.0	2.4

TABLE 5
MEANS OF EXPERIMENTAL AND CONTROL GROUPS ON THE COOPERATIVE U.S. HISTORY TEST (STANDARD SCORES: $\bar{X} = 50$, $S = 10$)

	N	Mean Fall	Mean Spring	Gain
Experimental Teachers - Regular Curriculum - Year 1	140	44.1	49.3	5.2
Experimental Teachers - Modified Curriculum - Year 2	280	41.3	49.7	8.4
Control Teachers - Regular Curriculum - Year 1	36	44.1	47.6	3.5
Control Teachers - Regular Curriculum - Year 2	51	39.5	46.9	7.4

1. **I.D.S. Test.**

a. Considered as a group, students of the experimental teachers showed significantly greater gain ($p < .01$) the second year—i.e., under the modified curriculum—as compared to the previous year. The amount of the difference, when compared to available norms, indicates the improvement to be of practical importance. The students under the revised curriculum began the year with a mean score very near that typical of eleventh graders and, by the end of the year, scored at a level almost up to that of a sample of unselected college students and almost as high as previously reported groups in high school classes emphasizing critical thinking. Students of these teachers but without the revised curriculum showed the amount of gain to be expected during the course of a year. Both groups began the year with nearly identical mean scores.

b. The significant ($p < .05$) teacher-by-method interaction suggests that the curricular modifications are more effective with some teachers than with others.

c. When students experiencing the revised curriculum were compared with students in the regular curriculum (during the same year—different teachers), they showed significantly greater gain ($p < .01$). The gain for the students in the regular curriculum (two teachers) was almost identical for the two years.

It seems legitimate to conclude that the revised curriculum had a rather marked effect on critical thinking as measured by the I.D.S. Test.

2. **Watson-Glaser Test.**

a. The results for this test do not support the I.D.S. Test results. There is essentially no difference between the two groups of students taught by the experimental teachers in amount of gain. In both years, the gain is 2.8. The group experiencing the modified curriculum was slightly lower on the fall testing. For the first year group, the gain is from a percentile score of 77 to 83 while for the second year group (modified curriculum), the gain is from the 74th to the 81st percentile rank based on high school norms. Grade equivalent scores are not available for this test.

b. The comparison of experimental and control groups during the second year only is consistent with the foregoing analysis in showing no significant difference between the groups.

The results for this test provide no evidence for the modified curriculum. This finding is particularly disappointing in light of the fact that the Illinois study did find a significant superiority in amount of gain shown on this test by the students in the experimental group.

3. **Cooperative U.S. History Test.**

a. Students under the modified curriculum made significantly more gain during the year than did students with the same teachers during the preceding year ($p < .001$). In both instances, the students at the end of the year scored slightly below national norms. The experimental group, however, scored considerably lower at the beginning of the year.

b. The experimental group (modified curriculum) showed more gain than the control group during the same year, but not significantly so.

c. The control teachers achieved significantly ($p < .05$) more gain the second year.

d. The gain of the experimental teachers was not significantly greater than the gain achieved by the control teachers during the second year. Because of the gain achieved by the experimental teachers, we are tempted to suggest that the curricular modifications may have fostered greater interest and/or skill in dealing with the course content, hence, greater mastery. But since the gain was not significantly greater than that achieved by the control teachers during the second year, it is possible that other factors were operative, possibly that the second year students began the year with somewhat poorer background. It is clear that the modifications did not result in a decrease in the mastery of course content.

Reactions of Teachers, Students, and Parents. An additional measure of the outcomes of a plan such as this is to be found in the reactions of persons involved in it. Although no systematic attempt was made to collect such data in the present study, some information almost inevitably is present. It is recognized that impressions such as those which follow are subject to many criticisms on the grounds of selective sampling and bias of several kinds; they are nevertheless presented as valuable, though for the most part subjective, data.

1. The seven experimental teachers have all expressed considerable enthusiasm for the program as an interesting and worthwhile attempt in an important area, though some are quite skeptical as to the results achieved, particularly among the less able students. Even accounting for the expected desire to comfort the researchers and to justify their own efforts, it is our opinion that this represents an honest reaction on the part of the teachers. One bit of supportive data is that they have all indicated an intention to use at least part of the materials next year and have expressed the hope that further work of this kind will be undertaken.

The consensus seems to be that the material on fallacies and definitions was easiest to put across, with the material on syllogisms the most difficult, as would be expected. As to organization of presentation, some of the teachers indicated that they would prefer to spread the topics out during the year and introduce them as smaller units. One teacher would, in the future, not teach the material as a distinct unit but rather would attempt to incorporate it throughout the course.

2. As reported by the teachers, the reaction of students was varied. Some expressed the view that it was difficult. Others wondered what it was for, i.e., "Why don't we just have history?" Our expectation was that some students would be psychologically threatened by the material; this seems to have been the case but to a lesser extent that we expected. On the other hand some became intrigued and enjoyed it. Several teachers reported students making use of the material in arguments and particularly in debate,

though some of the same material frequently is presented in debate (and in psychology courses). Several incidents of carryover to other activities were reported:

a. Letters were written to several advertisers and to a weather man requesting definition of terms. The former were not satisfactorily answered; the latter was – and in some detail.

b. As a result of a difference of opinion in class regarding a syllogism, several students wrote to a professor of philosophy at the University of Utah for clarification.

3. There appears to have been little reaction from parents. As expected, some parents feared that knowledge of history was being sacrificed for some new silliness, but the teachers were able to provide an explanation which was at least in some cases considered adequate.

We had expected some objection from parents along the lines that their children were beginning to question some of the eternal verities. That this did not happen may be attributable to the parents' confidence in the schools, to parental indifference, or to lack of impact of our program.

Summary

This report describes a two year project which introduced into three high schools a curriculum plan designed to foster critical thinking and which attempted to assess its effectiveness. The curriculum plan was patterned after a similar program developed at the University of Illinois and consisted of the presentation of a three week unit on the tools of logical analysis, semantics, and scientific method at a level appropriate to eleventh graders, followed by application of these tools to the content of the course in U.S. History throughout the year. The seven participating teachers were provided a workshop prior to the introduction of the unit and were provided the services of the project staff, as well as the benefits of several group discussions during the year. Their interest and effort expended in the project was such as to leave no question but that the approach received an adequate trial.

The results of the evaluation demonstrate quite clearly that mastery of the typical content of the U.S. History course was not impaired by the curriculum modification. The effectiveness of the program in fostering critical thinking is not unequivocally demonstrated, since one of the tests to assess this change did not show any difference between experimental and control groups. The other test, however, which on logical grounds may be argued to be a better test, did show rather impressive differences in favor of students who received the revised curriculum. Further, the reactions of teachers and students, though not intensively studied, strongly support the value of the program.

References

1. Aschner, Mary J. "Teaching the Anatomy of Criticism," *The School Review* LXIV (1956), pp. 317-22.

2. Carpenter, Finley. "Educational Significance of Studies on the Relation Between Rigidity and Problem Solving," *Science Education* XL (1956), pp. 296-311.

3. Ennis, Robert H. "An Appraisal of the Watson-Glaser Critical Thinking Appraisal," *Journal of Educational Research* LII (1958), pp. 155-158.

4. Ennis, Robert H. "Interim Report: The Development of the I.D.S. Critical Thinking Test," Personal Communication.

5. Henderson, Kenneth B. "The Teaching of Critical Thinking," *Phi Delta Kappan* XXXIX (1958), pp. 280-282.

The Analysis

Type of Study—quasi-experimental. The researchers did have control over the treatment but did not use random assignment of either students or teachers to treatment groups.

Justification. The authors relied on the "current acceptance" of its importance to justify studying critical thinking. While this is often done, we believe a reader deserves a more thorough treatment, perhaps something like the following:

> Many respected thinkers, including Dewey, Adler, Toffler, and Taba, have defended the necessity of students' learning to be critical thinkers rather than passive channels for the transmission of information. The rate of information generation is such that no one can expect to master even a limited content area for more than a very short time. In academic areas, therefore, one must learn to evaluate new information and to see its relationship to previous knowledge. In the more general arena of daily life, the necessity for citizens of a democracy to sift and evaluate competing claims for their allegiance and endeavors has been recognized since the framing of the U.S. Constitution.

One might also expect some rationale for the teaching method involved. While implied in the existing report, a more explicit statement might be the following:

> If our definition of critical thinking is accepted, one teaching approach that is immediately suggested is direct instruction in the component skills (e.g., the recognition of logical fallacies). Each skill is presented to students in a manner commensurate with their level of knowledge; opportunities to practice the skills and receive feedback are provided.

Further, one might expect to find an exposition of the implications of study outcomes for theory and practice:

> If it is shown that the method is effective, additional support is provided for those wishing to disseminate it more widely. Teachers and others will have reason to expect that the desired outcomes will, in fact, occur. Further, such results would also support the general theory, espoused by Bruner and others, that high school students are capable of learning content customarily taught in college. Finally, additional evidence would exist to support the proposition that critical thinking can be taught in a straightforward manner to all high school students in much the same ways as other, more typical content, rather than depending on greater maturity or special talent on the part of students or teachers.

Lastly, the authors should have, at the outset, indicated that the study was a replication of other work and provided more details regarding the prior study. They could not have been expected to review all of the studies pertaining to critical thinking prior to that time, but some additional references would have been helpful. In reality, the systematic review of related literature is a distasteful task for many researchers—or so we believe—and hence is often done, as here, in a cursory fashion.

We see no reason for concern about the ethical implications of the study, though the authors did state, near the end of their report, that they had anticipated some objections from parents because students were being encouraged to question commonly held assumptions. Discussion of the philosophical/political ramifications of this issue is beyond the scope of a research report, but the authors might have explained why they had such expectations.

Clarity. The focus of the study seems clear—to obtain evidence of the extent to which the curriculum modifications improve critical thinking in high school students and affect acquisition of customary knowledge of history. The primary outcome variable, "ability to think critically," is clear at the outset. Other outcome variables, however, were not mentioned until near the end of the study. These variables—reactions of teachers, students, and parents—should have been mentioned in the introduction.

In any study involving a complex treatment such as this, it is virtually impossible to convey all of the intricacies of the method involved. In our opinion, the authors presented as good a description as could be expected.

Hypotheses were not stated explicitly. We would argue that they should have been, since the study was clearly intended to test the efficacy of a particular method. The following six hypotheses were clearly implied, however.

During the new curriculum year, as compared to the preceding year, the classes of the experimental teachers will demonstrate:

1. Greater gain on the Watson-Glaser Test.
2. Greater gain on the I.D.S. Test.
3. Approximately the same amount of gain on the Cooperative U.S. History Test.

During the same year, classes taught

the new curriculum (the "experimental group" teachers), as compared to classes taught the usual curriculum (the "control group" teachers), will demonstrate:

4. Greater gain on the Watson-Glaser Test.

5. Greater gain on the I.D.S. Test.

6. Approximately the same amount of gain on the Cooperative U.S. History Test.

Definitions. No specific section on definitions was provided. The authors did provide somewhat of a constitutive definition of *critical thinking*. However, the statement that various definitions encompass "some or all" of these features is imprecise. Did the authors intend to include all the features? If not, which ones? Further explication would have been helpful, especially of the "correct principles of logic." Additional clarity could easily have been achieved by defining *critical thinking* operationally as the scores on the Watson-Glaser and I.D.S. tests. The essentials of the curriculum modifications are probably clear "in context" later in the report, but might have been called to the reader's attention earlier. Items a-d on the bottom on page 38 might well have been given as the definition of *critical thinking*, since they are more specific both as to the intent of the curriculum and its content.

Sample. The sample was clearly not obtained in a random manner, including as it did a total of nine teachers in three high schools and a total of 27 intact classes of students, all in one particular city. The authors did not argue for representativeness, since they would have had to offer evidence that the teachers and students were similar to a population of interest in some important ways (e.g., ability level, socioeconomic level of the students, years of experience of the teachers). In fact, the mean scores on the Cooperative U.S. History Test (see page 42) suggest that the student sample was very similar to, but slightly below, the normative group for that test. The sample, then, was a convenience sample, with all of its inevitable limitations.

Whether or not the authors wanted to argue for the generalizability of their results, they should have provided some demographic data. For example, the ethnic makeup of both the teacher and student samples can be presumed (from the location of the study) to be predominantly Anglo (as, in fact, was the case). Further, many readers would likely infer (again because of location) that the attitudes of the teachers would be highly conservative (though this was *not* the case). Since these variables would be expected to influence outcomes, some information on them should have been given. While the sample of students is large in all comparisons of interest, the sample of teachers is not (only seven experimental and two control). While actually larger than in many studies of this type, this sample size – particularly that of the control group – presents further limitations.

Threats to Internal Validity

History. It is always conceivable that one or more other factors, instead of the independent variable, may be responsible for the outcome(s) of a study. In this study, such factors might have included the availability of additional resources to experimental classes but not to control classes, a schoolwide disruption (i.e., a teachers' strike) during year one of the study, and the introduction of critical thinking materials into the physical science curriculum during year two. In any study, one must rely on the integrity and acumen of the researchers to identify and discuss such factors. Since none were mentioned here, we can only infer that none were known to the researchers. The study design – comparing groups both across years and within the same year – is probably the best way to rule out such possibilities, since they would not have been expected to favor the "new curriculum" group under both circumstances.

Maturation of students would have affected all comparison groups in the same way, since the pre/post interval was the same for all. Maturation of teachers might have accounted for the superiority of year two over year one results if the teachers were relatively inexperienced (this was not the case), but would have been unlikely to have accounted for differences in year two alone.

Mortality in students would not have been expected to favor the new curriculum groups, since it occurred either by absence from class or by random deletion.

Subject characteristics are always of concern when random assignment of sizable numbers of subjects is not used. Analysis of covariance and similar techniques (e.g., analysis of regressed gain scores) do make it possible to match groups with respect to measured variables (in this case, pretest scores), but cannot ensure comparability on other variables, such as student attitude toward social studies or interest in analytic processes. Further, such analyses make mathematical assumptions (such as how to determine the "best fit" line), which are themselves subject to sampling error.

In this study, the researchers should have identified those subject variables that were likely (1) to affect the outcome variable(s) and (2) to be different for the comparison group. They should then have attempted to measure these variables and incorporate them into the analysis. That this is easier said than done is illustrated by the difficulty of get-

ting socioeconomic data (probably one of the most important variables to control).

In a methods study such as this, researchers must also be concerned about possible differences between teachers of the two groups – perhaps the experimental teachers were just better teachers than the control teachers. Use of the same teachers for both methods – as in part of this design – is the best way to control for this threat.

Pretesting should not have given an advantage to the new curriculum group, since it was done in all groups. One might argue that the pretest interacted with the method to result in an advantage to the experimental group, but this seems unlikely, in that the pretest items were only a sample of the tasks emphasized all year long. Omission of the pretest would have eliminated this possibility at the sacrifice of statistical matching of groups.

A regression effect is unlikely, since extreme groups were not used. If anything, such an effect would favor the control group during year two, since it had lower pretest scores.

An instrumentation effect resulting in bias seems unlikely, since instrumentation was the same for all groups. It is conceivable that the new curriculum students might do more poorly on the posttest because of increased critical ability, but this would be contrary to the hypotheses and the outcomes obtained. It seems unlikely that bias would be introduced by test scoring, since all tests were machine-scored. Information on test administration should have been included, however. Administration of tests by teachers is notorious for violations of standard testing procedures. Had this been the case, one might suspect the experimental teachers of giving assistance or additional time in taking the tests. This, of course, would have favored their students. In actuality, this threat was eliminated, since project-trained assistants administered all tests.

A Hawthorne effect was a major concern in this study. Since both teachers and students knew that they were part of a special project and since the experimental teachers received special summer training, it could be argued that this special attention accounted for the results obtained. The only way to control for this threat would have been to provide similar special attention to the control groups.

An order effect would not apply to the students in this study, but it might be thought to have affected teachers, since they were involved during successive years. It seems unlikely, however, that second-year gains were larger because of first-year participation, since the first year consisted only of testing and organizational meetings.

The authors of the study did pay some attention to these threats, although in a rather general way. They did state that the design of the study permitted analysis of gains made from year to year by the same teacher and from teacher to teacher within a given year, but they did not indicate what specific threats were addresssed by this design. They also described how analysis of covariance matched the groups on pretest scores, but again they did not indicate what threat this controlled. Beyond these statements, there was no discussion of internal validity.

As to whether the treatment received an adequate trial, no in-class observation was reported. The statement that the project staff met periodically with the teachers throughout the year, however, combined with the examples of assignments developed and of teacher and student reactions lead us to conclude that the tryout was adequate in terms of substance. One year appears to be ample time for implementation of the curriculum.

Reliability of Instruments. The authors did a poor job of addressing reliability. They are guilty of the typical "quick shuffle" in stating that usage and "other evidence" were sufficient. They should, at the very least, have reviewed previous evidence as to type of reliability and the magnitude of reliability coefficients and then assessed their applicability to this study. Since the student sample appeared to be quite similar in performance on these tests to available norm groups, prior data might have been applicable. However, there is still no excuse for not reporting internal consistency coefficients, since they could easily have been obtained from the data available. While pre/post correlations are somewhat misleading as indicators of reliability in a treatment study (since one expects inconsistency pre to post), they are nevertheless of interest, particularly for comparison among the groups. If the new curriculum turned out to be effective, one might expect less pre/post consistency for students exposed to this curriculum, since new treatments are, by their nature, trying to disturb the predictable pattern of development.

Validity of Instruments. The authors provided a brief logical analysis of the two critical thinking tests. They did not, however, discuss these tests in relation to the curriculum modifications introduced in the study. Readers can make their own comparisons between the five subtests of the Watson-Glaser test and the outline of curriculum topics, but they should not have to do so. It appears that all five subtests have logical relevance to the curriculum topics, but that two topics (definitions and reliability of sources) may not have been tested. The authors had a responsibility to defend their use of this test as it relates to the content taught.

Even less information was provided on the validity of the I.D.S. test. While the use of inde-

pendent judges to assess the validity of these tests for the purposes of this study may be less crucial than in many studies reviewed in this monograph, it would have greatly strengthened the authors' report.

Finally, the authors neglected to report a very useful piece of information. They had a built-in empirical check on validity—the correlation between the two tests—which could easily have been obtained from the data at hand. It would be very helpful to have this correlation (both pre and post) separately for each major treatment group. The results of the group comparisons did suggest that these correlations were not high, but the details are important.

External Validity

Population Generalizability. To their credit, the authors did not specifically overgeneralize their results to "teachers" and "students," but rather phrased both their discussion of results and their summary in terms of the outcomes obtained for the teachers and students involved in the study. They failed to discuss the serious limitations imposed by their convenience sample, however. Also, their use of inferential statistics without qualification implies, we believe, that they thought their results were generalizable. They did mention that the study was a partial replication and that the replicated data did not support the previous findings, but we would argue that they should have included a statement somewhat like the following at the end of their summary:

> In total, our evidence indicates that further use and study of this method are warranted. We found no evidence of negative effects and some evidence of positive impact. Since, however, our results were equivocal and, in specifics, inconsistent with a prior study, and since our sample does not permit generalization to a defined population, these results must be treated as tentative.

Ecological Generalizability. The authors made no comments about the ecological generalizability of this study. They did not commit the (not uncommon) error of recommending this method in all social studies courses or at a variety of grade levels or in the absence of a university support system, but neither did they warn against such overgeneralization.

Results and Interpretations. The authors generally maintained a clear distinction between results and interpretations. When presenting the results for the U.S. history test, an interpretation was made, but it is clear that it was an inference going beyond the data. One might, however, question the inclusion of the section on reactions of teachers, students, and parents in the "Results" section, since some of the interpretations made there are not supported by the data presented. Overall, the data did, in our opinion, justify the interpretations made.

Data Analysis

Descriptive Statistics. The statistics presented are appropriate, although the omission of standard deviations is unfortunate. Standard deviations are important, in that they permit the assessment of the magnitude of the differences in mean gain. This is particularly important when inferential statistics are suspect (see below). Effect size also should have been reported.

Another method of judging the magnitude of change is by comparison with other groups. As the authors pointed out, the year one and year two experimental teacher groups began at very near the expected mean. While the year one group gained somewhat more than might be expected from normative data, the additional gain of the year two group (to that attained by "special groups") does seem sufficient to warrant the conclusions drawn.

More serious is the failure of the authors to explore and discuss the teacher-by-method interaction. This is a good example of both a preoccupation with hypothesis-testing and the political implications of research. The researchers felt justified in not reporting interaction data, since it was not primary to the purpose and implied hypotheses of the study. In actuality, however, means were examined and the results found to be somewhat disconcerting. The teacher who had been most enthusiastic and hardworking with respect to the new curriculum obtained the lowest gains during year two, while the teacher judged to have the poorest grasp of the content showed the highest gains. These findings were not only repugnant to the researchers, but also potentially destructive to the morale of the teachers in the experimental group. Consequently, nothing was said. We believe it is not uncommon in educational research to ignore awkward results because they are inconsistent with the current presumed state of knowledge and/or have implications that are potentially detrimental to interpersonal relationships. This situation is exacerbated by the all-too-common practice of terminating inquiry into a topic when a research grant runs out. In this instance, follow-up examination of this unexpected and distasteful outcome might have led to some truly significant findings. With the hindsight of twenty years, it now appears that the explanation for this result could be found in the quality of interpersonal relations between teacher

and students. Unfortunately, no data were collected on this variable.

Inferential Statistics. Analysis of covariance is an appropriate procedure for this study. Since the assumption of random sampling was violated, however, the authors were obligated to indicate that the resulting probabilities were not exact and should be interpreted only as general indications. This they did not do. It is legitimate to use the probabilities as indicators of greater gain on the I.D.S. test than on the Watson-Glaser, although the comparison of means (see Tables 3 and 4) makes the same point. What is not defensible (although common) was the reporting and interpreting of probabilities as though they could be taken at face value.

Significance of the Study. Despite the many criticisms we have made of the study, we judge it to be significant. This judgment reflects our appraisal of the importance of the topic and the realization that no study is perfect. Nevertheless, our analysis reveals that even experienced researchers can substantially improve the quality of their research and the reporting of their findings.

CHAPTER 5
SUGGESTIONS FOR CLASSROOM RESEARCH IN THE SOCIAL STUDIES

The profession has largely overlooked one group of individuals as not only a potential source of valuable information about social studies, but also as potential gatherers of such information. We refer to those most intimately involved with social studies subject matter, methodology, and classrooms – classroom teachers of social studies.

It appears to be a fact that most classroom teachers of the social studies do not engage in research. We found no reports of research efforts by classroom teachers in TRSE or the JSSR for the past ten years; only an occasional article by teachers can be found in the research section of SE during this same period. Similarly, recent reviews of research in social studies education reveal few studies by classroom teachers (e.g., Hunkins et al. 1977; Stanley 1985; Armento 1986).

Although it is only logical to assume that most social studies teachers want to improve the quality of their instructional efforts and thus probably experiment with new materials and approaches from time to time, there seems to be little desire on their part to engage in systematic research in their classrooms or to consider research as a source of ideas about possible ways to improve their efforts. We think this is unfortunate.

This is not, of course, a new idea (see, for example, Shaver 1979b; Wiley 1977). The intent of this section is therefore not to analyze in depth why social studies teachers neither engage in nor read research.[1] Let us just state briefly that they are not trained to do so in their social studies or general methodology courses; they are not encouraged to do so by their supervisors or administrators; they are not in any way rewarded for doing so; and they are further discouraged from such activity by the large numbers of students (often between 30 and 40 students per class) that most must teach. Even those few who read the research literature rarely find anything that, in their perception, relates directly to what they do in their own classrooms.[2]

Classroom teachers could investigate many kinds of questions in social studies education. Indeed, by doing so they could perform a vital service to the profession. It is an unfortunate fact that we still have only the haziest of ideas as to what sorts of content, methods, learning activities, teaching strategies, and evaluation devices make much, if any, sort of a difference in social studies classrooms; how students "learn" social studies most effectively; what methods work best in what sorts of situations; how to encourage and develop student thinking about social issues; how to vary content, methods, and activities to help students of differing abilities; how best to sequence content so as to maximize understanding; or even (alas!) how to increase the interest of the vast majority of students in the social studies curriculum itself.[3]

Classroom teachers can help to provide some answers to these (and other) important questions. In fact, if several teachers, in different schools within districts and even in different districts throughout the nation, were to investigate the same question in their classrooms, thereby replicating the research of their peers, they could begin to establish what might become a steadily accumulating base of knowledge about important aspects of teaching and learning in the social studies. As we indicated earlier, such a knowledge base, though badly needed, does not at present exist.

We believe there is another important reason why teachers and other school personnel might profitably conduct research – as a means of reducing burnout. In our experience, many teachers find it difficult to maintain enthusiasm for their work after several years of coping with all the stresses of the profession. Participation in research to clarify questions of interest and concern might be one of the best ways to maintain the intellectual excitement that, for many, has been lost.

In this chapter, therefore, we suggest methodologies classroom teachers could use to investigate questions of interest and then describe how a classroom teacher might use them to investigate one or more questions of interest. The techniques we suggest are designed not to be too demanding of their time and energy. The methodologies hold promise for providing information of interest and value not only to individual teachers but to the profession as a whole.

In the examples that follow, we use the discipline of history as the source for the research questions we present. Similar examples could be presented using other disciplines (political science, economics, etc.) or such topics as global education or law studies, which typically borrow information from a variety of disciplines.

Experimental Research

Suppose that a history teacher is interested in the following question: "How can I most effectively teach historical concepts to my students?" The

teacher might compare the effectiveness of certain methods of instruction (e.g., inquiry, case studies, illustrated lectures, programmed units, small group discussions) with others in promoting the learning of historical concepts. If conditions permit adequate controls, *experimental* research would be an appropriate methodology. Students could be systematically assigned to contrasting forms of history instruction. The effects of these contrasting methods could then be compared by testing the conceptual knowledge of those taught. Student learning could be assesed by an objective test, with the validity of the test checked in some way. The scores on the test (the dependent, or outcome, variable), if they differ, would give us some idea of the effectiveness of the two methods.

In the simplest sort of experiment where there are two contrasting methods to be compared (usually referred to as the independent variable), an attempt is made to control for all other (extraneous) variables, such as student ability level, age, grade level, time, materials, teacher characteristics, etc., that might affect the outcome under investigation. Methods of control could include randomly assigning students to one or the other of the instructional groups, holding the classes during the same or closely related periods of time, using the same materials in both groups, comparing students of the same age and grade level, etc.[4]

If possible, of course, one wants to have as much control as possible over the assignment of individuals to the various treatment groups. As we mentioned in Chapter 3, however, random assignment of students to treatment groups is usually difficult, if not impossible, to achieve. Nevertheless, comparisons are still possible. For example, achievement in two or more *intact* history classes in the same school, taught by teachers whose methods differ rather dramatically (predominantly lecture-oriented vs. discussion-oriented teachers, for example), might be compared. Since the students in these classes would not have been assigned to their classes randomly, this could not be considered a "true" experiment. Large differences between the classes, however, could still be suggestive of how the two methods compare.[5] Furthermore, it might be possible to compare groups that are matched on important variables—at least on a pretest.

Consider for a moment the study we analyzed in Chapter 4. With certain modifications, such a study could be carried out by any interested classroom teacher. Granted, the curriculum modifications were complex and required training, but any method a teacher wished to study could be substituted. A minimum of two classes is required; secondary teachers often teach several sections of the same course. Elementary teachers would need to divide their class randomly, compare succeeding classes (by semesters or years), or involve a second teacher. Obtaining the tests used, or others, should not be a problem. Data collection is a simple matter. Data analysis using gain scores (from pre to post) for each student is a straightforward and relatively simple process involving only means, medians, standard deviations, and frequency polygons. We believe that the mechanics of carrying out such a study, therefore, are by no means prohibitive.

The difficult, but also interesting, part is attending to the various issues we have discussed herein so as to arrive at legitimate and useful interpretations. Such efforts might well make truly significant contributions to the education of children and would go a long way toward reprofessionalizing teaching. We are well aware of the potential for erroneous conclusions on the part of individual teachers, but we believe that this can be counteracted by the insistence that intentions, plans, methods, and findings be shared with colleagues. Good research procedures, once demystified, are well within the grasp of most school personnel. Lastly, we think the probable benefits of our recommendations far outweigh any potential for misunderstanding.

Survey Research

Another teacher might not be interested in comparing instructional methods. He or she might say, "I'm more interested in the general feelings my students have about history. What do they like about their history classes? What do they dislike? Why? What types of history are liked the best or least? How do the feelings of students of different ages, sexes, and ethnicity in our school compare? In our district?"

These sorts of questions can best be answered through a variety of *survey* techniques that measure student attitudes toward their history classes. Questionnaires or interview schedules would need to be prepared and their validity and reliability checked in some way; the instruments could then be given to students, teachers, counselors, or other appropriate individuals to complete.

The difficulties involved in survey research are mainly twofold: (1) insuring that the questions to be answered are clear and not misleading (this can be accomplished, to a fair extent, by using objective or "closed-ended" questions, insuring that they all pertain to the topic under investigation, and then further eliminating ambiguity by a small pilot testing of a draft of the questionnaire); and (2) getting a sufficient number of the questionnaires completed and returned from the intended group so that meaningful analyses can be made (this can be

furthered by a second, and sometimes third, administration of the questionnaire to non-returnees). The big advantage of questionnaire research is that it can provide a lot of information from quite a large sample of individuals. If more details about particular questions are desired, however, a teacher can also conduct personal interviews with students. The advantage of an interview (over a questionnaire) is that open-ended questions (i.e., those requiring a written response of some length) can be used with greater confidence, particular questions of special interest or value can be pursued in depth, follow-up questions can be asked, items that are unclear can be explained, etc. Care must be taken not to forget, however, that data obtained through surveys is only a description of what *is* and not necessarily what should be. Survey results can, however, suggest possible hypotheses to investigate using some of the other methods described in this chapter.[6]

Content Analysis

Yet another teacher might be interested in the accuracy of the images or conceptions presented to students in their history textbooks. She or he might ask, "Is the content (written or visual) presented to students in history texts biased in any way, and if so, how?" Answering this question calls for a *content analysis*. A content analysis is just what its name implies—an analysis of the written or visual contents of a document. A person's or group's conscious and unconscious beliefs, attitudes, values, and ideas are often revealed in the things they write (draw, paint, etc.) in magazines, newpapers, novels, plays, advertisements, and books. Since history textbooks are comprised primarily of written material, this material can be analyzed in any one of a number of ways. To analyze the contents of a textbook (or textbooks), however, a teacher first needs to plan how to select and order the contents that are available for analysis. Pertinent categories must be developed so the teacher can identify and then count and compare that which he or she thinks is important.

This is the nub of content analysis—defining as precisely as possible those aspects of the content the teacher wants to investigate and then formulating relevant categories that are so explicit that another teacher who uses them to examine the same material would find essentially the same proportion of topics emphasized or ignored.

Suppose, for example, that a teacher is interested in the sorts of heroes being presented to students in various history textbooks. He or she would first select the sample of texts to be analyzed—that is, which textbooks he or she would read, on what subject(s), covering what time period, and which editions (e.g., current U.S. history texts available for use in his or her district). Categories could then be formulated. Possibilities might include the physical, emotional, and social characteristics of heroes; these could in turn be broken down into even smaller coding units, such as the following:

Physical	Emotional	Social
hair color	warm	race
eye color	aloof	religion
age	hostile	occupation
etc.	etc.	etc.

A coding sheet would then be prepared to tally the data in each of the categories as it is identified in each text. Comparisons could then readily be made.

A major advantage of content analysis is that it is unobtrusive. The teacher can "observe" without being observed, since the "contents" being analyzed are not influenced by the teacher's presence. Information that might be difficult or even impossible to obtain through direct observation or other means can be gained through analysis of textbooks and other available communication material without the author or publisher realizing that it is being examined. Furthermore, replication of a content analysis by another teacher is relatively easy. Thirdly, the information obtained through a content analysis of textbooks can be very helpful in planning for instruction. Such information can suggest additional data students need to get a more accurate and complete picture of the world in which they live, the factors and forces within it, and how these factors and forces impinge on people's lives.[7]

Correlational Research

A teacher might ask, "How can we predict which sorts of individuals are likely to have trouble learning historical subject matter?" If we could make fairly accurate predictions in this regard, then perhaps we could suggest some corrective measures teachers could employ to help such individuals, avoiding production of "history-haters." In this instance, *correlational research* may be the most appropriate methodology. An interested teacher could use a variety of measures to collect different sorts of data on students, including their performance on a number of tasks related to history learning (e.g., reading historical accounts, utilizing maps), their demographic characteristics, aspects of their backgrounds, their early experiences with history courses and history teachers, the kinds of history courses they have taken, and anything else that might conceivably point up how those students who do well (learn history) differ from those who do poorly.

The teacher might then look for *patterns* of some sort in each group of students (those who learn easily and those who have difficulty). What do those who learn history easily seem to have in common? What do they seem to be doing that those who have trouble learning history seem to ignore or avoid? What do they apparently *not* do?

The information obtained from such research can help a teacher predict more accurately the likelihood of learning difficulties for certain types of students in history courses and even, perhaps, suggest some things to try with different groups of students to help them learn.[8]

In short, correlational research seeks to investigate whether one or more relationships of some type exist. The approach requires no manipulation or intervention on the part of the teacher other than that required to administer the instrument(s) necessary to collect the data desired. In general, this type of research would be undertaken when a teacher wants to look for and describe relationships that may exist among naturally occurring phenomena, without trying in any way to alter these phenomena.

Causal-Comparative Research

A teacher might pursue additional sorts of investigations in which the variables involved cannot be manipulated. For example, a teacher[9] might be interested in discovering whether high school students enrolled in a college-bound curriculum feel differently about history than high school students enrolled in a non-college-bound curriculum. If this question were to be investigated experimentally, students would have to be randomly assigned into college-bound and non-college-bound curricula, and then their attitudes compared by means of one or more assessment devices. Conceptually, of course, this is possible but actually it would be impossible to do.

To test this question using a causal-comparative design, however, two groups of students who *already* exist, one enrolled in a college-bound program and the second in a non-college-bound program, could be compared to see if they differ in their feelings toward history. Suppose they do. Can the teacher conclude that the difference in curricula produced the difference in feelings? Alas, no. The teacher can only conclude that a relationship on some sort exists, but he or she cannot say what caused the relationship.

Thus, interpretations of causal-comparative research are limited because the teacher cannot say whether a particular variable is a cause or a result of the behavior(s) observed. In the example presented here, he or she would not know if any perceived differences in feelings between the two groups were due to the enrollment in the different type of curricula; if enrollment in the different curricula was due to a difference in attitude between the two groups; or if some third, unidentified, factor was at work.

Despite problems of interpretation, causal-comparative studies are of value in identifying *possible* causes of observed variations in the behavior patterns of students. These possible causes can then be investigated using experimental or other methods of research. Furthermore, additional information can sometimes strengthen the argument for causation, as in the research linking cigarette smoking and cancer.

Ethnographic and Case Study Research

In all of the examples so far presented, the questions being asked involve *how well* or *how much* or *how accurately* social studies learnings or attitudes or ideas exist or are being developed. Thus, possible avenues of research include comparisons between alternative methods of teaching social studies (using history as an example), surveying different groups of social studies students or social studies professionals (teachers, supervisors, etc.), or analyzing different social studies texts.

Quite another type of question can be asked about the teaching and learning of social studies, however. A teacher might be interested in knowing *not* how much or how well or how accurately, but simply "how." In the case of history, just how do history teachers teach their subject? What kinds of things do they do as they go about their daily routine? What sorts of things do students do? In what kinds of activities do they engage? What explicit and implicit rules of the game in history classes seem to help or hinder the process of learning?

To gain some insight into these concerns, an *ethnographic* methodology can be utilized. A teacher who wishes to further his or her understanding of how history is actually taught would try to document or portray the everyday experiences of students (and teachers, if possible) in history classrooms. The focus would be on only one student or one classroom (or a small number of them at most). The teacher would observe the student or the classroom on as regular a basis as possible (perhaps during preparation period) and attempt to describe, as fully and as richly as possible, what he or she sees going on. Descriptions (a better word might be "portrayals") might depict the social atmosphere of the classroom; the intellectual and emotional experiences of students; the manner in which the teacher (student) acts toward and reacts to (other) students of different ethnicities, sexes, or abilities; how the "rules" of the classroom are learned, modified, and enforced; the kinds of ques-

tions the teacher (and students) ask; and so forth. The data to be collected might include detailed prose descriptions written on legal-sized tablets by the teacher/observer, audiotapes of pupil-student conferences, videotapes of classroom discussions, examples of teacher lesson plans and student work, sociograms depicting "power" relationships in a classroom, and flowcharts illustrating the direction and frequency of certain types of comments (e.g., the kinds of questions asked by teacher and students to one another and the responses different kinds produce).[10]

Ethnographic or case study research also lends itself well to a detailed study of individuals. Sometimes much can be learned from studying just one individual. For example, some students learn history very easily. In hopes of gaining insight into why this is the case, a teacher might observe one such student on a regular basis to see if there are any noticeable patterns of regularities in the student's behavior. The student's teachers (counselors, coaches, etc.), as well as the student, might be interviewed in depth. A similar series of observations and interviews might be conducted with a student who finds history very difficult to learn. As much information as possible (study style, attitudes toward history, approach to the subject, behavior in class, etc.) would be collected. Through the study of a somewhat unique individual, insights might be gained that would help the teacher with similar students in the future.

In short, then, the goal of a teacher engaging in ethnographic or case study research is to "paint a portrait" of a history (or any social studies) classroom (or an individual) in as thorough, accurate, and vivid a manner as possible so that others can also "see" that classroom, its participants, and what they do. Indeed, ethnographic research seems a particularly viable approach for use in classrooms. Teachers contemplating using this methodology should keep in mind the cautions and recommendations we made in Chapter 3 and consult one or more of the sources we mentioned on page 29.

Classroom teachers can (and should, we would argue) participate in this research endeavor. There is so much in our field about which we know so little. So many questions remain unanswered. So much information is needed.

A Final Word

We recognize that our suggestions do not easily fit into the typical daily activities of most teachers (or other personnel). We acknowledge also that carrying out such efforts requires additional time and energy, which—given the demands on teachers—may seem excessive. However, we know of teachers who have, despite their obligations, found it possible to carry out such studies—including quasi-experimental research. They tell us that the effort required was more than compensated for by the information gained and the intellectual stimulation provided by the process. Thus, we are encouraged to commend such endeavors to others.

Notes

1. For a classroom teacher's analysis of why so few of her peers do research (but also why she thinks they should), see McKee (1986).

2. For some further thoughts as to why classroom teachers do not engage in research, see McPhie (1979).

3. For some thoughts and data on student interest in the social studies, see Schug, et al. (1984) and Shaughnessy and Haladyna (1985). For a response to Schug, et al., see Allen (1984).

4. A basic, but clear discussion of experimental research in the classroom can be found in Ferguson (1986). Examples of experimental research in social studies education include Gilmore and McKinney (1986); Kleg, Karabinus, and Carter (1986); Yoho (1986); and Betres, Zajano, and Gumieniak (1984).

5. An extremely thorough treatment of this type of research can be found in Cook and Campbell (1979). Examples of quasi-experimental research in social studies education include Beem and Brugman (1986); Barnes and Curlette (1985); Hahn and Avery (1985); and McKenzie and Sawyer (1986). For more details and examples of how to do quasi-experimental research in social studies classrooms, see Shaver (1979a).

6. Some helpful ideas about survey research can be found in Smith (1986). Examples of questionnaire- or interview-type survey research in social studies education include Bennett (1984); Jantz, et. al. (1985); LeSourd (1984); and Schug and Birkey (1985).

7. A good beginning reference for content analysis research is Wiseman and Aron (1970). Examples of content analysis research in social studies education include Anyon (1978); Stanley (1984); Hahn and Blankenship (1983); Romanish (1983), and Saltonstall (1979).

8. Examples of correlational research in social studies education include Curtis (1983) and Haladyna, Shaughnessy, and Redsun (1982).

9. Although we continue to refer to only one teacher in these research examples, we would like to stress that more than one teacher might be involved in a research endeavor in social studies education. Two or more teachers, acting as a re-

search *team*, for example, might decide to conduct research in their classrooms or with students.

10. A clearly written introduction to ethnographic research for social studies teachers can be found in White (1986). Examples of case study or ethnographic research in social studies education include Adler (1984); Diem (1986); Goodman and Adler (1985); and Levstik (1986).

CHAPTER 6
STUDIES REVIEWED

1. Adler, S. (1984). A field study of selected student teacher perspectives toward social studies. *TRSE* 12 (1), pp. 13-30.

2. Allen, W.H., and VanSickle, R.L. (1984). Learning teams and low achievers. *SE* 48 (1), pp. 60-64.

3. Alverman, D.E., Boothby, P.R., and Wolfe, J. The effect of graphic organizer instruction on fourth graders' comprehension of social studies text. *JSSR* 8 (1), pp. 13-21.

4. Ball, D.W., Doss, A.R., and Dewalt, M.W. (1986). Level of teacher objectives and their classroom tests: Match or mismatch. *JSSR* 10 (2), pp. 27-30.

5. Barnes, B.R. and Curlette, W.L. (1985). Effects of instruction on teachers' global mindedness and patriotism. *TRSE* 13 (1), pp. 43-49.

6. Barnes, B.R., Stallings, W., and Rivner, R. (1981). Are the critics right about MACOS? *TRSE* 9 (1), pp. 35-44.

7. Barth, J.L. (1985). Egyptian social studies teachers' responses to the Barth/Shermis social studies preference scale. *JSSR* 9 (2), pp. 15-25.

8. Barth, J.L., and Sommersdorf, D.L. (1981). A study of preservice teachers' change in attitude measured by Barth/Shermis social studies preference scale. *JSSR* 5 (2), pp. 64-68.

9. Barth, J.L., and Shermis, S.S. (1980). A study of student teachers' effectiveness in applying inquiry questioning skills. *JSSR* 4 (2), pp. 34-37.

10. Beem, A.L., and Brugman, D. (1986). The effects of values development lessons on pupils' well-being, pleasure in school, mutual relationships and on pupils' valuational behavior during classroom dialogues. *TRSE* 14 (2), pp. 97-112.

11. Beck, W.W. (1979). Utilizing an interdisciplinary "teacher centered" model in developing individual social studies materials. *JSSR* 3 (1), pp. 1-7.

12. Bennett, C. (1984). Interracial contact experience and attrition among black undergraduates at a predominantly white university. *TRSE* 12 (2), pp. 19-48.

13. Bennett, C.T. (1980). Citizenship traditions in the commonwealth of Virginia. *JSSR* 4 (1), pp. 17-21.

14. Bennett, C.T. (1981). What should be the educational objectives in the old dominion? *JSSR* 5 (1), pp. 46-48.

15. Betres, J., Zajani, M., and Gumieniak, P. (1984). Cognitive style, teacher methods, and concept attainment in social studies. *TRSE* 12 (2), pp. 1-18.

16. Blaga, J.L., and Nielsen, L.E. (1983). The status of state history instruction. *JSSR* 7 (1), pp. 45-57.

17. Blahut, J.M., and Nicely, R.F. (1984). Tactile activities and learning attitudes. *SE* 48 (2), pp. 153-155; 158.

18. Blankenship, G., Jr. (1983). The use of visual diagrams with eighth grade social studies students at four ability levels. *JSSR* 7 (1), pp. 23-44.

19. Boag, N. and Massey, D. (1981). Teacher perspectives on program change. *TRSE* 9 (3), pp. 37-60.

20. Burts, D.C., McKinney, C. W., and Burts, B.L. (1985). Effects of teacher enthusiasm on three- and four-year-old children's acquisition of four concepts. *TRSE* 13 (1), pp. 19-30.

21. Carney, J.J., Anderson, D., Blackburn, D., and Blessing, D. (1984). Preteaching vocabulary and the comprehension of social studies materials by elementary school children. *SE* 48 (3), pp. 195-196.

22. Chapin, J.R. (1982). Japanese and American youth's knowledge and attitudes on energy. *JSSR* 6 (1), pp. 17-21.

23. Curtis, C.K. (1981). Slow learners' attitudes toward fundamental freedoms. *JSSR* 5 (1), pp. 35-39.

24. Curtis, C.K. (1983). Relationships among certain citizenship variables. *JSSR* 7 (2), pp. 18-28.

25. Curtis, C.K., and Shaver, J.P. (1980). Slow learners and the study of contemporary problems. *SE* 44 (4), pp. 302-309.

26. Dally, A.R. (1983). Educational attainment and political attitudes: An effect of schools or schooling? *TRSE* 11 (2), pp. 35-52.

27. Diem, R.A. (1982). Measurement of social studies content knowledge in pre-service elementary education majors. *JSSR* 6 (1), pp. 8-12.

28. Diem, R.A. (1986). Computers in a school environment: Preliminary report of the social consequences. *TRSE* 14 (2), pp. 163-170.

29. Eddinger, S.S. (1985). The effect of different question sequences on achievement in high school social studies. *JSSR* 9 (1), pp. 17-29.

30. Eyler, J. (1980). Citizenship education for conflict: An empirical assessment of the relationship between principled thinking and tolerance for conflict and diversity. *TRSE* 8 (2), pp. 11-26.

31. Eyler, J. (1982). A test of a model relating political attitudes to participation in high school activities. *TRSE* 10 (1), pp. 43-62.

32. Farmer, R. (1983). Elementary social studies teacher education: Some points to ponder. *JSSR* 7 (2), pp. 29-36.

33. Ferguson, P. (1986). The effects of teaching microcomputer programming skills to social studies teachers. *JSSR* 10 (1), pp. 53-56.

34. Fleming, D.B., and Weber, L.J. (1979). Recognizing point of view: A critical reading skill in the social studies. *SE* 44 (2), pp. 153-156.

35. Ford, M.J., and McKinney, C.W. (1986). The effects of recycling and response sensitivity on the acquisition of social studies concepts. *TRSE* 14 (1), pp. 21-34.

36. Gilmore, A.G., and McKinney, C.W. (1986). The effects of student questions and teacher questions on concept acquisition. *TRSE* 14 (3), pp. 225-244.

37. Glenn, A.D., and Ellis, A.K. (1982). Direct and indirect methods of teaching problem solving to elementary school children. *SE* 46 (2), pp. 134-136.

38. Goetz, J. P. (1981). Children's sex-role knowledge and behavior: An ethnographic study of first graders in the rural south. *TRSE* 8 (4), pp. 31-54.

39. Goodman, J., and Adler, S. (1985). Becoming an elementary social studies teacher: A study in perspectives. *TRSE* 13 (2), pp. 1-20.

40. Grant, E.T., and Napier, J.D. (1981). Comparing an aesthetic and a political approach to teaching world history. *SE* 45 (5), pp. 372-375.

41. Gunn, T.Y. (1980). The relationship between students' ability to read a textbook and achievement. *JSSR* 4 (1), pp. 22-25.

42. Guskey, T.R. (1981). The relationship of affect toward teaching and teaching self-concept to responsibility for student achievement. *JSSR* 5 (2), pp. 69-74.

43. Hahn, D.L. (1985). The diffusion of an innovation: A case study of one social studies program. *JSSR* 9 (2), pp. 26-39.

44. Hahn, C.L. (1986). Are teachers prejudiced against students writing on non-traditional topics for their gender? *JSSR* 10 (1), pp. 31-39.

45. Hahn, C.L., and Avery, P.G. (1985). Effect of value analysis discussions on students' political attitudes and reading comprehension. *TRSE* 13 (2), pp. 47-60.

46. Haladyna, T., Shaughnessy, J., and Redsun, A. (1982a). Correlates of attitudes toward social studies. *TRSE* 10 (1), pp. 1-26.

47. Haladyna, T., Shaughnessy, J., and Redsun, A. (1982b). Relations of student, teacher, and learning environment variables to attitudes toward social studies. *JSSR* 6 (2), pp. 36-44.

48. Hamilton, S.F. (1981). Adolescents in community settings: What is to be learned? *TRSE* 9 (2), pp. 23-38.

49. Hastings, W.L. (1986). Political socialization themes in the post-Watergate era. *SE* 50 (6), pp. 453-457.

50. Jackstadt, S.L., and Brennan, J.M. Economic knowledge and high school student attitudes toward the American economic system, business, and labor unions. *TRSE* 11 (3), pp. 1-16.

51. Jacko, C.M., Karmos, A.H., and Karmos, J.S. (1982). What classroom teachers think and do about sex-role stereotyping. *JSSR* 6 (1), pp. 32-34.

52. Jantz, R.K. (1981). Children's attitudes toward the elderly: A look at Greek and American children. *TRSE* 9 (2), pp. 1-22.

53. Jantz, R.K., Weaver, V.P., Cirrincione, J.M., and Farrell, R.T. (1985). Inquiry and curriculum change: Perceptions of school and college/university faculty. *TRSE* 13 (2), pp. 61-72.

54. Jeter, J.T., and Davis, O.L., Jr. (1982). Differential classroom interaction in social studies as a function of differential expectations of pupil achievement. *JSSR* 6 (1), pp. 1-7.

55. Kennedy, K.J. (1983). Assessing the relationship between information processing capacity and historical understanding. *TRSE* 11 (2), pp. 1-22.

56. King, N.R. (1982). School uses of materials traditionally associated with children's play. *TRSE* 10 (3), pp. 17-28.

57. Kingston, P.W., and Bennett, C.T. (1986). Improving high school social studies: Advice from the colleges. *TRSE* 14 (1), pp. 35-50.

58. Kirman, J.M. (1984). A new elementary level map skill: Landsat "Band 5" satellite images. *SE* 48 (3), pp. 191-194.

59. Kleg, M., Karabinus, A., and Carter, T. (1986). Direct concept instruction in U.S. history. *JSSR* 10 (2), pp. 1-12.

60. Krall, F., and Holt, L. (1980). A comparison of "openness" of student teachers in alternative and typical classrooms. *JSSR* 4 (1), pp. 13-16.

61. Land, M.L., and Smith, L.R. (1981). College students' ratings and teacher behavior: An experimental study. *JSSR* 5 (1), pp. 19-22.

62. Larkins, A.G, and McKinney, C.W. (1982). Two studies of the effects of teacher enthusiasm on the social studies achievement of seventh grade students. *TRSE* 10 (1), pp. 27-41.

63. LeSourd, S.J. (1984). An exploratory comparison of two methods for assessing teacher attitudes toward instructional strategies. *TRSE* 12 (1), pp. 31-41.

64. LeSourd, S.J. (1985). Using text structure to improve social science concept attainment. *JSSR* 9 (2), pp. 1-14.

65. Levstik, L.S. (1986). The relationship between historical response and narrative in a sixth-grade classroom. *TRSE* 14 (1), pp. 1-20.

66. Long, S. (1980). Urban adolescents and the political system: Dimensions of disaffection. *TRSE* 8 (1), pp. 31-44.

67. Maddocks, W.H., and Smith, L.R. The effects of student IQ and programmed instruction frame length on social studies achievement. *JSSR* 10 (1), pp. 40-51.

68. Marker, G.W. (1980). Why schools abandon "new social studies" materials. *TRSE* 7 (4), pp. 35-57.

69. Masters, P.E., Jr. (1984). Political socialization in an international city: The case of Atlanta. *JSSR* 8 (2), pp. 17-38.

70. Martin, W.C. (1981). Transferring the learning of teaching models. *JSSR* 5 (1), pp. 40-45.

71. McCutcheon, G. (1981). Elementary school teachers' planning for social studies and other subjects. *TRSE* 9 (1), pp. 45-66.

72. McGowan, T.M. (1984). Does methodology make a difference? A comparison of instructional practices of teacher and student attitudes toward social studies. *JSSR* 8 (1), pp. 22-39.

73. McKenzie, G.R., and Sawyer, J. (1986). Effects of test-like practice and mnemonics on learning geographic facts. *TRSE* 14 (3), pp. 201-210.

74. McKinney, C.W. (1985). A comparison of the effects of a definition, examples and nonexamples on student acquisition of the concept of "transfer propaganda." *SE* 49 (1), pp. 66-70.

75. McKinney, C.W., Ford, M.J., and McKinney, J.C. (1984). The effect of presentation order of examples and nonexamples on undergraduates' aquisition of a social studies concept. *JSSR* 8 (1), pp. 1-12.

76. McKinney, C.W., and Larkins, A.G. (1982). Effects of high, normal, and low teacher enthusiasm on secondary social studies achievement. *SE* 46 (4), pp. 290-292.

77. McKinney, C.W., Larkins, A.G., and Burts, D.C. (1984). Effects of overt teacher enthusiasm on first-grade students' acquisition of three concepts. *TRSE* 11 (4), pp. 15-24.

78. McKinney, C.W., Larkins, A.G., and Peddicord, H.Q., Jr. (1982). A comparison of three instructional designs for teaching social studies concepts to fourth grade students. *JSSR* 6 (2), pp. 45-47.

79. McTeer, J.H. (1979). Sex differences in students' interest in certain discipline areas of the social studies. *JSSR* 3 (2), pp. 58-64.

80. Moore, J.R., and Williams, P.L. (1980). Trends in social studies graduation competencies. *TRSE* 8 (2), pp. 27-36.

81. Napier, J.D., and VanSickle, R.L. (1981). Evaluation of a peer microteaching and field experience component in a secondary level social studies preservice teacher education program. *JSSR* 5 (1), pp. 1-8.

82. Nelson, J.L., and Palonsky, S. (1980). Preservice teacher perceptions of social education. *JSSR* 4 (1), pp. 5-12.

83. Nelson, M.R. (1984). Social educators and social action. *JSSR* 8 (1), pp. 54-63.

84. O'Brien, L.M., Meszaros, B., and Pulliam, W.E. Effects of teachers' use of objectives on student achievement in social studies. *TRSE* 13 (3), pp. 57-65.

85. Palonsky, S., and Nelson, J. (1980). Political restraint in the socialization of student teachers. *TRSE* 7 (4), pp. 19-34.

86. Peterson, E.L, and Nett, E. M. (1982). Sexuality in the classrooms of teachers with various sex role orientations. *TRSE* 10 (3), pp. 29-40.

87. Powell, P.M., and Powell, J.V. (1984). An investigation of political apathy among selected high school students. *JSSR* 8 (2), pp. 53-66.

88. Reilly, M.E. (1979). Eliminating sexism: A challenge to educators. *SE* 43 (4), pp. 312-316.

89. Richardson, H.W. (1979). Communication, leadership, and decision making: A comparative study of three supervisors. *JSSR* 3 (1), pp. 8-15.

90. Ritchie, C.A., and Steinbrink, J.E. (1984). Knowledge, attitude, and the energy ethic. *SE* 48 (2), pp. 149-153.

91. Robinson, P. (1980). The status of the secondary social studies methods course in Arkansas. *JSSR* 4 (2), pp. 38-42.

92. Russell, T.E., and Morrow, J.E. (1986). Reform in teacher education: Perceptions of secondary social studies teachers. *TRSE* 14 (4), pp. 325-330.

93. Schober, H. (1984). An analysis of the impact of teacher training in economics. *TRSE* 12 (1), pp. 1-12.

94. Schug, M.C. (1983). Elementary teachers' views on economic issues. *TRSE* 11 (1), pp. 55-66.

95. Schug, M.C., and Birkey, C.J. (1985). The development of children's economic reasoning. *TRSE* 13 (1), pp. 31-42.

96. Schug, M.C., Todd, R.J., and Beery, R. (1984). Why kids don't like social studies. *SE* 47 (5), pp. 382-387.

97. Schultz, W. (1979). An experiment using selected sociological concepts to affect attitudinal change. *JSSR* 3 (2), pp. 38-42.

98. Shermis, S.S., and Wasburn, P.C. (1986). Social studies educators and their beliefs: Preliminary data from Indiana colleges and universities. *TRSE* 14 (4), pp. 331-340.

99. Slack, J.P., and Larkins, A.G. (1982). The effect of two instructional treatments on college students' map skills achievement. *JSSR* 6 (1), pp. 13-16.

100. Smith, B.D. (1980). Influence of solicitation pattern, type of practice example, and student response on pupil behavior, commitment to discussion, and concept attainment. *TRSE* 7 (4), pp. 1-18.

101. Smith, B.D. (1983). Instructional planning: Attitudes, decisions, and preparation time among secondary social studies teachers. *JSSR* 7 (1), pp. 1-22.

82

102. Smith, L.R. (1984). Lesson commonality and method of reading effect on achievement in social studies. *JSSR* 8 (2), pp. 1-16.

103. Smith, L.R. (1985). The effect of teacher uncertainty and student ability level on achievement in social studies. *JSSR* 9 (1), pp. 30-40.

104. Sorgman, M., Krall, F., and Uhlenberg, D. (1980). Interests of upper elementary students in human geography and their perceptions of effective learning environments. *JSSR* 4 (1), pp. 1-4.

105. Stahl, R.J. (1981). Achieving values and content objectives simultaneously within subject matter-oriented social studies classrooms. *SE* 45 (7), pp. 580-585.

106. Stahl, R. J., and Matiya, J.C. Teaching psychology in the high school: Does area of certification translate into different types of teachers and courses? *TRSE* 9 (2), pp. 55-86.

107. Stanley, W.B., Charlesworth, R., and Ringuest, J.L. (1985). Kindergarten and first grade children's social concept development. *JSSR* 9 (1), pp. 11-16.

108. Sunal, C.S., Paul, M.S., and Demary, J. (1981). Social studies for the hearing impaired: The state of the art. *TRSE* 9 (3), pp. 61-72.

109. Switzer, T.J., Walker, E., and Mitchell, G. (1981). Undergraduate social studies methods instructors' knowledge and use of new curricular methods. *JSSR* 5 (1), pp. 9-18.

110. Tlou, J.S., and Bennett, C.T. (1982). Social studies educators' attitudes toward censorship in public schools in the commonwealth of Virginia. *JSSR* 6 (2), pp. 48-52.

111. Tlou, J., and Bennett, C. (1983). Teacher perceptions of discipline problems in a central Virginia middle school. *JSSR* 7(2), pp. 37-59.

112. Turner, T.N. (1982). Controlling research with human subjects and its effects on educational research: Report of a survey of what teacher educators think. *JSSR* 6 (1), pp. 29-31.

113. Tyo, J. (1980). An alternative for poor readers in social science. *SE* 44 (4), pp. 309-310.

114. VanSickle, R.L., and Ehman, L.H. (1981). Developing and evaluating teacher training materials: A case study. *JSSR* 5 (2), pp. 52-63.

115. Walstad, W.B., and McFarland, M. (1980). "Trade-offs" and teacher in-service. *SE* 44 (5), pp. 410-411.

116. Wilson, A.H. (1979). Experience and social studies education. *JSSR* 3 (2), pp. 43-49.

117. Wilson, A.H. (1981). Where have all the graduates gone: A follow-up survey of 26 social studies education graduates. *JSSR* 5 (1), pp. 23-34.

118. Yoho, R.F. (1986). Effectiveness of four concept teaching strategies on social studies concept acquisition and retention. *TRSE* 14 (3), pp. 211-224.

REFERENCES

Adler, S. (1984). A field study of selected student teacher perspectives toward social studies. *Theory and Research in Social Education* 12 (1), pp. 13-30.

Agar, M.H. (1986). *Speaking of Ethnography*. Beverly Hills, CA: Sage.

Allen R.F. (1984). Why kids really don't like social studies. (Letter to the Editor). *Social Education* 48 (7), p. 498.

Allender, J.S. (1986). Educational research: A personal and social process. *Review of Educational Research* 56 (2), pp. 173-194.

Anyon, J. (1978). Elementary social studies textbooks and legitimating knowledge. *Theory and Research in Social Education* 6(3), pp. 40-55.

Armento, B.J. (1986). Research on teaching social studies. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 942-951). New York: Macmillan.

Babbie, E. (1986). *The Practice of Social Research* (4th ed.). Belmont, CA: Wadsworth.

Bakan, D. (1975). Speculation in psychology. *Journal of Humanistic Psychology*, 15 (1), pp. 17-25.

Barnes, B.R., and Curlette, W.L. (1985). Effects of instruction on teachers' global mindedness and patriotism. *Theory and Research in Social Education* 13 (1), pp. 43-49.

Barr, R.D., Barth, J.L., and Shermis, S. (1977). *Defining the Social Studies* (Bulletin 51). Arlington, VA: National Council for the Social Studies.

Barzun, J., and Graff, H.F. (1977). *The Modern Researcher* (3rd ed.). New York: Harcourt, Brace, Jovanovich.

Beem, A.L., and Brugman, D. (1986). The effects of values development lessons on pupils' well-being, pleasure in school, mutual relationships and on pupils' valuational behavior during classroom dialogues. *Theory and Research in Social Education* 14 (2), pp. 97-112.

Bennett, C. (1984). Interracial contact experience and attrition among black undergraduates at a predominantly white university. *Theory and Research in Social Education* 12 (2), pp. 19-48.

Berlak, A., and Berlak, H. (1981). *Dilemmas of Schooling: Teaching and Social Change*. New York: Methuen.

Betres, J., Zajano, M., and Gumieniak, P. (1984). Cognitive style, teacher methods, and concept attainment in social studies. *Theory and Research in Social Education*, 12 (1), pp. 1-18.

Bogdan, R.C., and Biklen, S.K. (1982). *Qualitative Research for Education: An Introduction to Theory and Methods*. Boston: Allyn and Bacon.

Borg, W.R., and Gall, M.D. (1983). *Educational Research: An Introduction* (4th ed.). New York: Longman.

Campbell, D.T., and Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of Research on Teaching* (1st ed., pp. 171- 246). Chicago: Rand-McNally.

Carr, E.H. (1967). *What Is History?* New York: Random House.

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review* 48, pp. 378-399.

Cherryholmes, C. (1982). Discourse and criticism in the social studies classroom. *Theory and Research in Social Education* 9 (4), pp. 57-74.

Common, D.L. (1982). Small group instruction in social studies classrooms and the corruption of critical thought. *Theory and Research in Social Education* 10 (4), pp. 49-68.

Cook, T.D., and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Studies*. Boston: Houghton-Mifflin.

Cornbleth, C. (1982). On the social study of social studies. *Theory and Research in Social Education* 10 (4), pp. 1-16.

Cornbleth, C. (1985). Reconsidering social studies curriculum. *Theory and Research in Social Education* 9 (4), pp. 57-74.

Curtis, C.K. (1983). Relationships among certain citizenship variables. *Journal of Social Studies Research* 7 (2), pp. 18-28.

Deutscher, I. (1973). *What We Say/What We Do*. Glenview, IL: Scott, Foresman.

Diem, R.A. (1986). Computers in a school environment: Preliminary report of the social consequences. *Theory and Research in Social Education* 14 (2), pp. 163-170.

Dobbert, M.L. (1982). *Ethnographic Research: Theory and Application for Modern Schools and Societies*. New York: Praeger.

Egan, K. (1980). John Dewey and the social studies curriculum. *Theory and Research in Social Education* 8 (2), pp. 37-55.

Ehman, L.H., and Hahn, C.L. (1981). Contributions of research to social studies education. In H.D. Mehlinger and O.L. Davis, Jr. (Eds.), *The Social Studies*. Eightieth Yearbook of the National Society for the Study of Education, Part II (pp. 60-81). Chicago: National Society for the Study of Education.

Eisner, E.W. (1981). On the differences between scientific and artistic approaches to qualitative research. *Educational Researcher* 10 (4), pp. 5-9.

Eisner, E. W. (1983). Anastasia might still be alive, but the monarchy is dead. *Educational Researcher* 12, pp. 13-14; 23-24.

Ferguson, P. (1986). Experimental research. In C. Cornbleth (Ed.), *An Invitation to Research in Social Education*. Washington, D.C.: National Council for the Social Studies.

Fraenkel, J.R. (1987). Toward improving research in social studies education. *Theory and Research in Social Education* 15 (3), pp. 203-222.

Gilmore, A.G., and McKinney, C.W. (1986). The effects of student questions and teacher questions on concept acquisition. *Theory and Research in Social Education* 14 (3), pp. 225-244.

Giroux, H.A., and Penna, A.N. (1979). Social education in the classroom: The dynamics of the hidden curriculum. *Theory and Research in Social Education* 7 (1), pp. 21-42.

Goodman, J., and Adler, S. (1985). Becoming an elementary social studies teacher: A study in perspectives. *Theory and Research in Social Education* 13 (2), pp. 1-20.

Gordon, B. (1985). Toward emancipation in citizenship education: The case of African-American cultural knowledge. *Theory and Research in Social Education* 12 (4), pp. 1-24.

Gottschalk, L. (1969). *Understanding History: A Primer of Historical Method*. New York: Alfred A. Knopf.

Hahn, C.L., and Avery, P.G. (1985). Effect of value analysis discussions on students' political attitudes and reading comprehension. *Theory and Research in Social Education* 13 (2), pp. 47-60.

Hahn, C.L, and Blankenship, G. (1983). Women and economic textbooks. *Theory and Research in Social Education* 11 (3), pp. 67-76.

Haladyna, T., Shaughnessy, J., and Redsun, A. (1982). Relations of student, teacher, and learning environment variables to attitudes toward social studies. *Journal of Social Studies Research* 6 (2), pp. 36-44.

Hepburn, M.A., and Dahler, A. (1985). An overview of social studies dissertations, 1977-1982. *Theory and Research in Social Education* 13 (2), pp. 73-82.

Holmes, M.M. (1982). The new middle class and the organization of curricular knowledge. *Theory and Research in Social Education* 10 (2), pp. 33-44.

Hunkins, F.P., Ehman, L.H., Hahn, C.L., Martorella, P.H., and Tucker, J.L. (1977). *Review of Research in Social Studies Education: 1970-1975*. Arlington, VA: National Council for the Social Studies.

Hurst, J.B. (1980). Political pablum: Democratic role models in children's picture books. *Theory and Research in Social Education* 7 (3), pp. 1-20.

Jantz, R.K., Weaver, V.P., Cirrincione, J.M., and Farrell, R.T. (1985). Inquiry and curriculum change: Perceptions of school and college/university faculty. *Theory and Research in Social Education* 13 (2), pp. 61-72.

Kerlinger, F.M. (1986). *Foundations of Behavioral Research* (4th ed.). New York: Holt, Rinehart, and Winston.

Kinsey, A.C., Pomeroy, B., and Dartin, C.E. (1948). *Sexual Behavior in the Human Male*. Philadelphia, PA: W.B. Saunders Co.

Kleg, M., Karabinus, R., and Carter, T. (1986). Direct concept instruction in U.S. history. *Journal of Social Studies Research* 10 (2), pp. 1-12.

Larkins, A.G., and McKinney, C.W. (1980). Four types of theory: Implications for research in social education. *Theory and Research in Social Education* 8 (1), pp. 9-18.

Larkins, A.G., and McKinney, C.W. (1982). Two studies of the effects of teacher enthusiasm on the social studies achievement of seventh grade students. *Theory and Research in Social Education* 10 (1), pp. 27-41.

Larkins, A.G., and Puckett, J.L. (1983). Issues in a relatively nonrationalized study: Suggestions for interpretive research. *Theory and Research in Social Education* 11 (2), pp. 23-24.

Leming, J.S. (1985). Research on social studies curriculum and instruction: Interventions and outcomes in the socio-moral domain. In W.B. Stanley (Ed.), *Review of Research in Social Studies Education: 1976-1983*. Washington, DC: National Council for the Social Studies.

LeSourd, S.J. (1984). An exploratory comparison of two methods for assessing teacher attitudes toward instructional strategies. *Theory and Research in Social Education* 12 (1), pp. 31-41.

Levstik, L.S. (1986). The relationship between historical response and narrative in a sixth-grade classroom. *Theory and Research in Social Education* 14 (1), pp. 1-20.

Martorella, P.H. (1977). Research on social studies learning and instruction: Cognition. In F.P. Hunkins, L.H. Ehman, C.L Hahn, P.H. Martorella, and J.L. Tucker, *Review of Research in Social Studies Education: 1970-1975*. Washington, DC: National Council for the Social Studies.

McKee, S.J. (1986). A teacher's perspective. In C. Cornbleth (Ed.), *An Invitation to Research in Social Education*. Washington, DC: National Council for the Social Studies, pp. 119-126.

McKenzie, G.R., and Sawyer, J. (1986). Effects of test-like practice and mnemonics on learning geographic facts. *Theory and Research in Social Education* 14 (3), pp. 201-210.

McPhie, W.E. (1979). Should social studies teachers do research? *Social Education* 43 (7), pp. 587-588.

Metcalf, L.E. (1963). Research on teaching the social studies. In N.L. Gage (Ed.), *Handbook of Research on Teaching* (pp. 929- 965). Chicago: Rand McNally.

Mook, D.G. (1983). In defense of external invalidity. *American Psychologist* 38, pp. 379-387.

Morrissett, I., and Haas, J.D. (1982). Rationales, goals and objectives in social studies. In Project SPAN, *The Current State of Social Studies*. Boulder, CO: Social Science Education Consortium.

Nelson, J.L., and Shaver, J.P. (1985). On research in social education. In W.B. Stanley (Ed.), *Review of Research in Social Studies Education: 1976-1983*. Washington, DC: National Council for the Social Studies.

Newmann, F.M. (1985). The radical perspective on social studies: A synthesis and critique. *Theory and Research in Social Education* 13 (1), pp. 1-18.

Newmann, F.M. (1986). Priorities for the future: Toward a common agenda. *Social Education* 50 (4), pp. 240-250.

Popkewitz, T.S., Tabachnick, R., and Wehlage, G. (1981). *The Myth of Educational Reform*. Madison, WI: University of Wisconsin Press.

Powell, P.M., and Powell, J.V. (1984). An investigation of political apathy among selected high school students. *Journal of Social Studies Research* 8 (2), pp. 53-66.

Romanish, B.A. (1983). Modern secondary economics textbooks and ideological bias. *Theory and Research in Social Education* 11 (1), pp. 1-24.

Saltonstall, C.S. (1979). Explanations of poverty: a critical review of secondary level social studies textbooks in the United States. (Doctoral dissertation, Harvard University, 1978). *Dissertation Abstracts International* 39, 4034-A.

Schug, M.C., Todd, R.J., and Beery, R. (1984). Why kids don't like social studies. *Social Education* 47 (5), pp. 382-387.

Schug, M.C., and Birkey, C.J. (1985). The development of children's economic reasoning. *Theory and Research in Social Education* 13 (1), pp. 31-42.

Shaughnessy, J.M., and Haladyna, T.M. (1985). Research on student attitude toward social studies. *Social Education* 48 (10), pp. 692-695.

Shaver, J.P. (1979a). Designing teacher-conducted research: Guidelines for classroom teachers. *Social Education* 42 (10), pp. 589-593.

Shaver, J.P. (1979b). The usefulness of educational research in curricular/instructional decision-making in social studies. *Theory and Research in Social Education* 7 (3), pp. 21-42.

Shaver, J.P., and Larkins, A.G. (1973). Research on teaching social studies. In R.M.W. Travers (Ed.), *Handbook of Research on Teaching* (2nd. ed., pp. 1243-1262). Chicago: Rand McNally.

Shaver, J.P., and Norton, R.S. (1980). Populations, samples, randomness, and replication in two social studies journals. *Theory and Research in Social Education* 8 (2), pp. 1-10.

Smith, B.D. (1980). Influence of solicitation pattern, type of practice example, and student response on pupil behavior, commitment to discussion, and concept attainment. *Theory and Research in Social Education* 7 (4), pp. 1-18.

Smith, D.L. (1986). Survey research. In C. Cornbleth (Ed.), *An Invitation to Research in Social Education*. Washington, DC: National Council for the Social Studies.

Spindler, G. (Ed.) (1982). *Doing the Ethnography of Schooling: Educational Anthropology in Action*. New York: Holt, Rinehart and Winston.

Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher* 7 (2), pp. 5-8.

Stanley, W.B. (1981). Toward a reconstruction of social education. *Theory and Research in Social Education* 9 (1), pp. 67-85.

Stanley, W.B. (1984). Approaches to teaching concepts and conceptualizing: An analysis of social studies textbooks. *Theory and Research in Social Education* 11 (4), pp. 1-14.

Stanley, W.B. (Ed.) (1985). *Review of Research in Social Studies Education: 1976-1983*. Washington, DC: National Council for the Social Studies.

Stouffer, S.S. (1950). Some observations on study design. *American Journal of Sociology* 55, pp. 355-361.

VanSickle, R.L. (1983). Statistical power and effect size in social education research. *Journal of Social Studies Research* 7 (2), pp. 1-17.

Vockell, E.L. (1983). *Educational Research*. New York: Macmillan.

Wallen, N.E. (1974). *Educational Research: A Guide to the Process*. Belmont, CA: Wadsworth.

Wallen, N.E. (1983, November). *Research in Social Studies: What is Needed?* Paper presented at the annual meeting of the National Council for the Social Studies. San Francisco, CA.

Wallen, N.E., Haubrich, W.F., and Reid, I.E. The outcomes of curriculum modifications designed to foster critical thinking. *Journal of Educational Research* 56 (7), pp. 529-534.

Wallen, N.E. and Fraenkel, J.R. (1988). An analysis of social studies research over an eight year period. *Theory and Research in Social Education* 16 (1), pp. 1-22.

Wasburn, P.C. (1986). The political role of the American school. *Theory and Research in Social Education* 13 (2), pp. 31-46.

Webster's New World Dictionary of the American Language (2nd coll. ed.) (1984). New York: Simon and Schuster.

White, J.J. (1986). Ethnographic research. In C. Cornbleth (Ed.), *An Invitation to Research in Social Education*. Washington, DC: National Council for the Social Studies.

Wiersma, W. (1987). *Research Methods in Education: An Introduction* (4th ed.). Itasca, IL: Peacock.

Wiley, K.B. (1977). *The Status of Precollege Science, Mathematics, and Social Science Education: 1955-1975. Vol. III: Social Science Education*. Washington, DC: U.S. Government Printing Office.

Wiseman, J.P., and Aron, M.S. (1970). *Field Projects for Sociology Students*. Cambridge, MA: Schenkman Publishing Co.

Yoho, R.F. (1986). Effectiveness of four concept teaching strategies on social studies concept acquisition and retention. *Theory and Research in Social Education* 14 (3), pp. 211-224.